

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Инженерная школа информационных технологий и робототехники
 Направление подготовки 15.04.04 Автоматизация технологических процессов и производств
 Отделение школы (НОЦ) Отделение автоматизации и робототехники

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Разработка алгоритма предиктивной диагностики технологического процесса производства БОПП-пленки

УДК 004.896:004.724.2

Студент

Группа	ФИО	Подпись	Дата
8ТМ91	Поданев Максим Николаевич		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОАР ИШИТР	Заревич Антон Иванович	к.т.н.		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШПИБ	Гончарова Наталья Александровна	к.э.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ООД ШПИБ	Сечин Андрей Александрович	к.т.н.		

По разделу на иностранном языке

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИЯ ШБИП	Сидоренко Татьяна Валерьевна	к.п.н.		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОАР ИШИТР	Гайворонский Сергей Анатольевич	к.т.н., доцент		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ООП

Код компетенции	Наименование компетенции
Универсальные компетенции	
УК(У)-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий
УК(У)-2	Способен управлять проектом на всех этапах его жизненного цикла
УК(У)-3	Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели
УК(У)-4	Способен применять современные коммуникативные технологии, в том числе на иностранном (-ых) языке (-ах), для академического и профессионального взаимодействия
УК(У)-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия
УК(У)-6	Способен определять и реализовывать приоритеты собственной
Общепрофессиональные компетенции	
ОПК(У)-1	Способен формулировать цели и задачи исследования, самостоятельно изучать научно-техническую документацию своей профессиональной деятельности
ОПК(У)-2	Способен определить математическую и техническую сущность задач и провести их качественно-количественный анализ
ОПК(У)-3	Способен на основании статистических методов участвовать в проведении корректирующих и превентивных мероприятий, направленных на улучшение качества, интерпретировать и представлять результаты
ОПК(У)-4	Способен анализировать полученные результаты измерений на основе их физической природы и принимать обоснованные решения в области профессиональной деятельности
Профессиональные компетенции выпускников	
ПК(У)-1	Обладает способностью разрабатывать технические задания на модернизацию и автоматизацию действующих производственных и технологических процессов и производств, технических средств и систем автоматизации, управления, контроля, диагностики и испытаний, новые виды продукции, автоматизированные и автоматические технологии ее производства, средства и системы автоматизации, управления процессами, жизненным циклом продукции и ее качеством;
ПК(У)-2	Обладает способностью проводить патентные исследования с целью обеспечения патентной чистоты и патентоспособности новых проектных решений и определения показателей технического уровня проектируемой продукции, автоматизированных и автоматических технологических процессов и производств, средств их технического и аппаратно-программного обеспечения;

ПК(У)-3	Обладает способностью: составлять описание принципов действия и конструкции устройств, проектируемых технических средств и систем автоматизации, управления, контроля, диагностики и испытаний технологических процессов и производств общепромышленного и специального назначения для различных отраслей национального хозяйства, проектировать их архитектурно-программные комплексы;
ПК(У)-4	Обладает способностью разрабатывать эскизные, технические и рабочие проекты автоматизированных и автоматических производств различного технологического и отраслевого назначения, технических средств и систем автоматизации управления, контроля, диагностики и испытаний, систем управления жизненным циклом продукции и ее качеством с использованием современных средств автоматизации проектирования, отечественного и зарубежного опыта разработки
ПК(У)-5	Обладает способностью разрабатывать функциональную, логическую и техническую организацию автоматизированных и автоматических производств, их элементов, технического, алгоритмического и программного обеспечения на базе современных методов, средств и технологий проектирования;

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа ИШИТР

Направление подготовки (специальность) 15.04.04 Автоматизация технологических процессов и производств

Отделение школы (НОЦ) Отделение автоматизации и робототехники

УТВЕРЖДАЮ:

Руководитель ООП

_____ Гайворонский С.А.
 (Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

В форме:

магистерской диссертации

(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)

Студенту:

Группа	ФИО
8ТМ91	Поданев Максим Николаевич

Тема работы:

Разработка алгоритма предиктивной диагностики технологического процесса
производства БОПП-пленки

Утверждена приказом директора (дата, номер)

Срок сдачи студентом выполненной работы:

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p>Исходные данные к работе <i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	<ol style="list-style-type: none"> 1. Техническая информация об АСУТП технологического процесса производства БОПП-пленки. 2. Регламент технологического процесса 3. Исторические данных (в виде временных рядов) технологического процесса.
<p>Перечень подлежащих исследованию, проектированию и разработке вопросов <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов.</i></p>	<ol style="list-style-type: none"> 1. Анализ технологий для проведения исследования; 2. Тест временных рядов на стационарность и автокорреляцию; 3. Анализ методов прогнозирования; 4. Анализ методов поиска аномалий; 5. Сравнительный анализ предсказательных

<i>подлежащих разработке; заключение по работе).</i>	<p>моделей;</p> <p>6. Сравнительный анализ моделей поиска аномалий</p> <p>7. Выполнение предпроектного анализа, планирование управления научно-техническим проектом, расчет бюджета исследования;</p> <p>8. Анализ производственной безопасности, экологической безопасности, а также безопасности в чрезвычайных ситуациях. Исследование правовых и организационных вопросов обеспечения безопасности при проведении исследований.</p>
<p>Перечень графического материала <i>(с точным указанием обязательных чертежей)</i></p>	<p>1. Схема информационных потоков;</p> <p>2. Графики прогнозных моделей и моделей поиска аномалий;</p> <p>3. Рисунки конфигурация программного обеспечения;</p> <p>4. Диаграмма Ганта;</p> <p>5. Диаграммы распределения классов в наборе данных</p>
<p>Консультанты по разделам выпускной квалификационной работы <i>(с указанием разделов)</i></p>	
<p>Раздел</p>	<p>Консультант</p>
<p>Финансовый менеджмент, ресурсоэффективность и ресурсосбережение</p>	<p>Гончарова Наталья Александровна</p>
<p>Социальная ответственность</p>	<p>Сечин Андрей Александрович</p>
<p>Раздел на иностранном языке</p>	<p>Сидоренко Татьяна Валерьевна</p>
<p>Названия разделов, которые должны быть написаны на русском и иностранном языках:</p>	
<p>Описание технологического процесса производства БОПП-пленки</p>	
<p>Оборудование ТДО</p>	
<p>Схема информационных потоков</p>	
<p>Разработка предиктивной модели</p>	

<p>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</p>	
--	--

Задание выдал руководитель:

<p>Должность</p>	<p>ФИО</p>	<p>Ученая степень,</p>	<p>Подпись</p>	<p>Дата</p>
------------------	------------	------------------------	----------------	-------------

		звание		
Доцент ОАР ИШИТР	Заревич Антон Иванович	к.т.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ТМ91	Поданев Максим Николаевич		

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа ИШИТР

Направление подготовки (специальность) 15.04.04 Автоматизация технологических процессов и производств

Отделение школы (НОЦ) Отделение автоматизации и робототехники

Период выполнения (осенний / весенний семестр 2020 /2021 учебного года)

Форма представления работы:

магистерская диссертация

(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	20
	Социальная ответственность	20
	Раздел на иностранном языке	60

СОСТАВИЛ:

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОАР ИШИТР	Заревич Антон Иванович	к.т.н.		

СОГЛАСОВАНО:

Руководитель ООП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОАР ИШИТР	Гайворонский Сергей Анатольевич	к.т.н., доцент		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа		ФИО	
8ТМ91		Поданев Максим Николаевич	
Школа	ИШИТР	Отделение школы (НОЦ)	ОАР
Уровень образования	Магистр	Направление/специальность	15.04.04 Автоматизация технологических процессов и производств

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Использовать действующие ценники и договорные цены на потребленные материальные и информационные ресурсы, а также указанную в МУ величину тарифа на эл. энергию
2. Нормы и нормативы расходования ресурсов	—
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Действующие ставки единого социального налога и НДС, ставка дисконтирования = 0,1 (см. МУ)

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Оценка коммерческого и инновационного потенциала НТИ	Дать характеристику существующих и потенциальных потребителей (покупателей) результатов ВКР, ожидаемых масштабов их использования
2. Разработка устава научно-технического проекта	Разработать проект такого устава в случае, если для реализации результатов ВКР необходимо создание отдельной организации или отдельного структурного подразделения внутри существующей организации
3. Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок	Построение плана-графика выполнения ВКР, составление соответствующей сметы затрат, расчет цены результата ВКР.
4. Определение ресурсной, финансовой, экономической эффективности	Оценка экономической эффективности использования результатов ВКР, характеристика других видов эффекта

Перечень графического материала (с точным указанием обязательных чертежей):

1. «Портрет» потребителя результатов НТИ
2. Сегментирование рынка
3. Оценка конкурентоспособности технических решений
4. Диаграмма FAST
5. Матрица SWOT
6. График проведения и бюджет НТИ - выполнить
7. Оценка ресурсной, финансовой и экономической эффективности НТИ - выполнить
8. Потенциальные риски

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШПИБ	Гончарова Н.А.	К.Э.Н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ТМ91	Поданев Максим Николаевич		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8ТМ91	Поданев Максим Николаевич

Школа	ИШИТР	Отделение	ОАР
Уровень образования	Магистр	Направление/специальность	Автоматизация технологических процессов и производств

Тема ВКР:

Разработка алгоритма предиктивной диагностики технологического процесса производства БОПП-пленки

Исходные данные к разделу «Социальная ответственность»:

1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения.	Объектом исследования являются исторические данные (представленные в виде временных рядов) двигателей установки продольного растяжения БОПП-пленки. Результатом работы является предиктивная модель двигателей, разработанная на языке Python.
---	--

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Правовые и организационные вопросы обеспечения безопасности: - специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; - организационные мероприятия при компоновке рабочей зоны.	Трудовой кодекс РФ и СанПиН 2.2.2/2.4.1340-03
2. Производственная безопасность 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	ГОСТ 12.1.005-88 ССБТ. Общие санитарно-гигиенические требования к воздуху рабочей зоны. ГОСТ Р 55710-2013 Освещение рабочих мест внутри зданий. Нормы и методы измерений. ГОСТ Р 54944-2012. Здания и сооружения. Методы измерения освещенности. ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности. ГОСТ 12.1.006-84 ССБТ. Электромагнитные поля радиочастот. Общие требования безопасности. ГОСТ 15484-81 Излучения ионизирующие и их измерения, термины и определения. ГОСТ 12.1.019-2017 ССБТ. Электробезопасность. Общие требования и

	номенклатура видов защиты. ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования.
3. Экологическая безопасность:	3.1 Рассмотрены негативно влияющие на экологию факторы при эксплуатации компьютера. 3.2 Решения по обеспечению экологической безопасности согласно нормативным документам.
4. Безопасность в чрезвычайных ситуациях:	Возможные ЧС на объекте: возгорание, взрыв, короткое замыкание. Наиболее распространённым типом ЧС является пожар.

Дата выдачи задания для раздела по линейному графику	
---	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ООД ИШБИП	Сечин Андрей Александрович	к.т.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ТМ91	Поданев Максим Николаевич		

Реферат

Пояснительная записка содержит 157 страниц машинописного текста, 21 таблиц, 94 рисунка, список использованных источников из 75 наименований и шести приложений.

Ключевые слова: временные ряды, анализ временных рядов, ARIMA, машинное обучение, нейронные сети, контрольные карты.

Цель работы: потенциальное уменьшение количества остановок производства БОПП-пленки и планирование мероприятий по обслуживанию оборудования за счет предиктивных моделей.

Объект исследования – исторические данные (в форме временных рядов) о технологическом процессе: значения датчиков, исполнительных механизмов и другие параметры технологической линии.

Степень внедрения: интеграция в имитационный технологический процесс.

Область применения: технологический процесс производства БОПП-пленки.

Фактором экономического эффекта, достигаемым в предложенной работе, является увеличение надежности технологической линии, планирование мероприятий по обслуживанию оборудования за счет предиктивных моделей.

Для выполнения работы использовались программные продукты: язык программирования Python, среда разработки Jupyter Notebook, MS Excel для хранения набора данных, OPC-сервер «Process Simulator», «UaExpert», база данных InfluxDB, платформа онлайн-мониторинга Grafana, TIA Portal.

Оглавление

Реферат.....	11
Обозначения и сокращения.....	16
Определения.....	17
Введение.....	19
Цели и задачи исследования.....	20
Обзор существующих решений.....	21
1 Описание технологического процесса производства БОПП-пленки.....	23
2 Оборудование ТДО.....	28
2.1 Контроллерное оборудование ТДО.....	28
2.2 Датчики ТДО.....	29
2.3 Двигатели ТДО.....	30
2.4 Ректифер ТДО.....	30
2.5 Control Unit ТДО.....	31
3 Схема информационных потоков ТДО.....	32
4 Разработка предиктивной модели.....	33
4.1 Основные шаги в разработке предиктивной модели.....	33
4.2 Сбор данных для предиктивной модели.....	33
4.3 Постановка задачи.....	34
4.4 Требования к временному ряду.....	34
4.5 Стек технологий для исследования.....	38
4.6 Метрики качества.....	40
4.7 Подготовка данных.....	41
4.8 Прогнозные модели.....	42
4.8.1 Методы скользящего среднего (moving average, MA).....	43
4.8.2 Методы экспоненциального сглаживания.....	44
4.8.2.1 Экспоненциальное сглаживание.....	44
4.8.2.2 Двойное экспоненциальное сглаживание.....	45
4.8.2.3 Тройное экспоненциальное сглаживание (Holt-Winters).....	47
4.8.3 Авторегрессионные модели.....	49
4.8.3.1 AR (autoregressive) модель.....	49
4.8.3.2 ARIMA (autoregressive integrated moving average) модель.....	50

4.8.3.3 SARIMA (seasonal autoregressive integrated moving average) модель...	52
4.8.4 Скрытая марковская модель (Hidden Markov model).....	53
4.8.5 Регрессионные модели.....	56
4.8.5.1 Теоретическая часть.....	56
4.8.5.2 Практическая часть.....	57
4.8.6 Модели на базе классификационно-регрессионных деревьев.....	60
4.8.6.1 CatBoost.....	60
4.8.6.2 XGBoost.....	62
4.8.6.3 LightGBM.....	64
4.8.6.4 Сравнение CatBoost, XGBoost, LightGBM.....	66
4.8.7 Нейросетевые модели.....	67
4.8.8 Готовые решения (Facebook Prophet, Amazon Gluon TS).....	71
4.8.8.1 Facebook Prophet.....	71
4.8.8.2 Amazon Gluon TS.....	72
4.8.9 Сравнение моделей прогнозирования.....	74
4.9 Детектирование аномалий.....	76
4.9.1 Статистические методы.....	77
4.9.1.1 Ящик с усами.....	77
4.9.1.2 Z-score.....	78
4.9.2 Контрольные карты.....	79
4.9.3 Методы, основанные на плотности.....	80
4.9.3.1 Локальный уровень выброса (LOF, Local outlier factor).....	80
4.9.3.2 DBSCAN.....	81
4.9.4 Методы машинного обучения.....	82
4.9.4.1 One Class SVM.....	82
4.9.4.2 Isolation Forest.....	84
4.9.5 Модельные методы.....	85
4.9.6 Методы на основе экспертных знаний.....	86
4.9.7 Сравнение методов детектирования.....	87
4.10 Интеграция предиктивной модели в технологический процесс производства БОПП-пленки.....	90
4.10.1 Имитационная модель ТДО.....	90

4.10.2 SCADA-пакет.....	91
4.10.3 OPC-сервер.....	92
4.10.4 База данных.....	92
4.10.5 Онлайн-мониторинг технологического процесса.....	93
4.10.6 Реализация взаимодействия технологического процесса и предиктивной модели.....	94
5 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение..	98
5.1 Организация и планирование работ.....	98
5.1.1 Продолжительность этапов работ.....	99
5.2 Расчет сметы затрат на выполнение проекта.....	107
5.2.1 Затраты на материалы и покупные изделия.....	107
5.2.2 Затраты на заработную плату.....	108
5.2.3 Затраты на социальный налог.....	109
5.2.4 Затраты на электроэнергию.....	109
5.2.5 Амортизационные расходы.....	110
5.2.6 Расходы, учитываемые непосредственно на основе платежных документов.....	111
5.2.7 Прочие расходы.....	112
5.2.8 Общая себестоимость разработки.....	112
5.2.9 Доход.....	112
5.2.10 Затраты на НДС.....	113
5.2.11 Цена разработки НИР.....	113
5.3 Оценка экономической эффективности проекта.....	113
6 Социальная ответственность.....	115
6.3 Правовые и организационные вопросы обеспечения безопасности.....	116
6.3.1 Особенности трудового законодательства.....	116
6.3.1 Особенности трудового законодательства.....	117
6.4 Производственная безопасность.....	118
6.5 Анализ опасных и вредных факторов.....	120
6.6 Экологическая безопасность.....	129
6.7 Безопасность в чрезвычайных ситуациях.....	130
Заключение.....	132
Список литературы.....	133

Приложение А (обязательное) Description of the manufacturing process of the BOPP film TDO equipment Scheme of information flows Development of a predictive model algorithm.....	140
Приложение Б.....	156
Приложение В.....	157
Приложение Г.....	159
Приложение Д.....	160
Приложение Е.....	161

Обозначения и сокращения

В данной работе применены следующие термины с соответствующими определениями:

АРМ – автоматизированное рабочее место;

АСУ – автоматизированная система управления;

АСУТП – автоматизированная система управления технологическим процессом;

БОПП – биаксиально-ориентированная полипропиленовая плёнка;

ДЭС – двойное экспоненциальное сглаживание;

МДО – машина продольной ориентации;

МНК – метод наименьших квадратов;

ПЛК – программируемый логический контроллер;
ТДО – установка поперечной ориентации;
ТОиР – техническое обслуживание и ремонт;
ТП – технологический процесс;
ТЭС – тройное экспоненциальное сглаживание;
СУУТП – система усовершенствованного управления технологическим процессом;
ЭС – экспоненциальное сглаживание;
AR – autoregressive;
ARMA – autoregressive moving-average model;
ARIMA – autoregressive integrated moving average;
CU – ток двигателя;
MA – moving average;
NM – момент двигателя;
SARIMA – seasonal autoregressive integrated moving average;
SVM – support vector machine;
SD – скорость двигателя;
TE – температура двигателя;
WMA – weighted moving average.

Определения

В данной работе применены следующие термины с соответствующими определениями

АСУ: Комплекс аппаратных и программных средств, а также персонала, предназначенный для управления различными процессами в рамках технологического процесса, производства, предприятия.

АСУТП: человеко-машинный комплекс, обеспечивающий управление технологическими процессами на современных механизированных и автоматизированных промышленных предприятиях.

База данных: набор перманентных (постоянно хранимых) данных, используемых прикладными программными системами какого-либо предприятия.

Временной ряд: ряд расположенных в хронологической последовательности значений статистического показателя, характеризующего изменения явления во времени.

Интерфейс: логическая и (или) физическая граница между устройством и средой передачи информации.

Технологический процесс: это последовательность связанных действий, которые выполняются с того момента, как возникли исходные данные и до требуемого результата.

Протокол: правила, управляющие обменом информацией. Определяет форму (семантику и синтаксис) сообщений.

ПЛК: это программно-управляемый дискретный автомат, имеющий множество входов, подключенных с помощью датчиков к объекту управления. Также имеющий множество выходов, подключенных к исполнительным устройствам.

Регрессия: задача прогнозирования метки с целевым значением для образца без метки.

Гиперпараметр: это свойство алгоритма обучения, обычно имеющее числовое значение.

Сезонность временного ряда: периодические повторяющиеся колебания анализируемого явления в определенное время года.

Цикл временного ряда: изменения изучаемого явления, обусловленные действием долговременных циклов.

Ошибка временного ряда: изменения не поддающиеся учету и регистрации.

Машинное обучение: раздел информатики, посвященный созданию алгоритмов, опирающихся на набор данных о каком-либо явлении.

Признак: значение, описывающее некоторую характеристику объекта.

СУБД: совокупность программных и языковых средств, предназначенных для управления данными в базе данных, ведения базы данных и обеспечения многопользовательского доступа к данным.

SCADA: аппаратно-программные компоненты АСУ ТП, которые обеспечивают взаимодействие пользователя с системой.

OPC-сервер: обеспечивает запись и чтение данных между клиентской программой и физическими устройствами.

Введение

Одна из главных тенденций современного производства – это концепция «Индустрия 4.0» [1]. «Индустрия 4.0» предполагает цифровизацию производства, внедрение в него киберфизических систем, предиктивной аналитики, использование виртуальной реальности для обучения сотрудников, работу с большими данными, создание цифровых двойников. Тенденция появилась потому, что крупные компании и России, и мира стремятся увеличить экономическую эффективность, повысить безопасность и надежность производства, автоматизировать процессы на всех уровнях. А одной из важнейших проблем, решаемой индустрией 4.0, является непрерывная работа оборудования.

На данный момент существует несколько методов обслуживания оборудования – аварийное, регламентное и предиктивное [2].

Аварийное обслуживание – обслуживание оборудования после выхода его из строя. Данный метод оправдан при обслуживании дешевого оборудования, не оказывающего решающего влияния на технологический процесс, либо имеющего резервирование.

Регламентное обслуживание осуществляется плановыми проверками согласно документации оборудования. План проверок рассчитывается с помощью статистического анализа компанией-изготовителем. Проблема возникает при работе оборудования на высоких нагрузках, что повышает шанс выхода из строя раньше плана проверки и соответственно приводит к нарушению технологического процесса.

Третий вид обслуживания – предиктивный. Когда процесс контролируется непрерывно по фактическому состоянию. Часто происходит так, что отклонения производственного процесса от обычного своего течения обнаруживается только по факту самого отклонения. Выявление таких отклонений (аномалий) происходит с помощью предсказательной аналитики, то есть строится модель предсказаний, имитирующая ход процесса и

прогнозирующая его на несколько шагов вперед, после чего возникает подзадача – детектирование аномалий в прогнозируемой части процесса. Данный метод позволяет сформировать мероприятия технического обслуживания к моменту отклонения процесса, что уменьшает количество фактических неисправностей, простоев, затрат на обслуживание и т.д.

Данная работа посвящена применению методов машинного обучения для предиктивной аналитики технологического процесса производства БОПП-пленки с последующим детектированием аномалий.

Производство БОПП-пленки – это сложный технологический процесс, в котором некорректная работа какого-либо оборудования может привести к большому времени простоев, что, в свою очередь, приведет к значительным экономическим потерям. Поэтому предсказание процесса и выявление в нём аномалий существенно увеличит экономическую эффективность производства, так как уменьшит вероятность нарушения хода работы технологической линии.

Цели и задачи исследования

Цель исследования – потенциальное уменьшение количества остановок производства БОПП-пленки и планирование мероприятий по обслуживанию оборудования за счет предиктивных моделей.

Объект исследования – исторические данные (в форме временных рядов) о технологическом процессе: значения датчиков, исполнительных механизмов и другие параметры технологической линии.

Предмет исследования – методы прогнозирования временного ряда и детектирования аномалий в нём.

Задачи исследования:

1. Сбор данных.
2. Подготовка данных.
3. Построение прогнозных моделей.
4. Детектирование аномалий.

5. Интеграция в технологический процесс.

Обзор существующих решений

В статье [3] исследуется прогнозирование отказов оборудования в условиях малого количества поломок. На основе алгоритма Random Forest строятся прогнозные модели, вычисляются отклонения прогнозных сигналов от фактического в каждый момент времени. В результате возникают проблемы с определением контрольных границ отклонения сигнала, а также с интеграцией модели в производство. Также в статье нет аргументов в пользу выбора алгоритма Random Forest.

В статье [4] выделяются плюсы и минусы предиктивной аналитики в контексте внедрения методики на производство. Предлагается решение, рассматривающее оценку определенных показателей – функции надежности, функции риска. И имеющее следующие функциональные возможности – мониторинг работы технологического процесса, планирование работ по устранению неисправностей, построение прогноза. Решение не рассматривает варианты интеграции в производство, не предлагает прогнозирующую модель и существует только на уровне концепций.

В статье [2] исследуется модель предиктивного обслуживания оборудования с применением беспроводных сенсорных сетей. Предлагается использовать вместо проводных датчиков – беспроводные. Передавать данные с них в «облако», после чего применять различные статистические методы для прогнозирования выхода какого-либо оборудования. Проблема рассматриваемого решения в том, что передача информации с использованием беспроводных датчиков небезопасна и ненадежна. Кроме того, внедрение в уже функционирующее производство проблематично из-за стоимости и различных протоколов связи.

В магистерских диссертациях [5], [6], [7] исследуется поиск аномалий во временных рядах. Предлагаются подходы для их выявления, сравнивается эффективность различных моделей. Но не анализируются реальные данные

производства из-за чего ценность работ в поиске оптимального метода для детектирования аномалий, а не в создании предиктивной модели.

В магистерской диссертации [8] применяются виртуальные агенты систем управления для обнаружения и ликвидации аварийных случаев. Предлагается внедрение системы ТОиР, основанной на агентных методах, которая повысит производительность производства и снизит риски. Предполагается, что система может быть применима практически к любой АСУ. Минусы данного решения – регламентированное обслуживание оборудования, что приводит к повышенным затратам предприятия.

В выпускной квалификационной работе [9] разрабатывается прогнозная модель изменения температуры. Для этого используется нейросетевой анализ. Решение не проводит сравнительный анализ прогнозных моделей, соответственно построение нейросетей возможно избыточно для данной задачи.

В бакалаврских работах [10], [11] производится анализ временных рядов. Рассматривается классификация моделей прогнозирования и временных рядов, алгоритмы построения прогнозных моделей. В работе [11] создается кроссплатформенное web-приложение в основе которого лежит модель прогнозирования фреймворка Fbprophet. Недостатками работы являются – использование базы данных SQLite, которая не предназначена для хранения временных рядов, следовательно при возникновении потребности масштабирования необходимо будет сменить базу данных. Кроме того, рассмотрены не все модели из классификации и не проведен сравнительный анализ их.

В статьях [12], [13] рассматриваются внедрение системы мониторинга с прогнозированием. В статье [12] основе модели прогнозирования лежит тройное экспоненциальное сглаживание, в статье [13] модель прогнозирования не строится. Системы мониторинга разрабатывались на языке Python, в качестве базы данных использовалась InfluxDB.

Недостатками работ являются отсутствие результатов внедрения, сравнительного анализа моделей и баз данных.

1 Описание технологического процесса производства БОПП-пленки

БОПП-пленка – биаксиально-ориентированная полипропиленовая пленка. Применяется для пищевой продукции и товаров повседневной жизни.

Технология производства БОПП-пленки начинается с приёма гранулированного сырья на участке разгрузки. Гранулят загружается по пневмотранспорту в силосы для хранения (рисунок 1.).



Рисунок 1 – Силос

Плёночное полотно состоит из 5-ти слоёв. Центральный слой формируется главным экструдером (рисунок 2.), куда поступает сырьё из силосов. Назначение главного экструдера – обеспечить за счет высоких температур и механического воздействия плавление материалов и гомогенизацию расплава по составу, температуре, давлению и вязкости. Пара вспомогательных сателлитных соэкструдера формируют промежуточные слои и служат для придания плёнкам дополнительных свойств в основном для внешнего вида. Крайние слои формируются двумя сателлитными экструдерами.



Рисунок 2 – Схема экструдера и соэкструдеров

Далее пленка поступает в свечевой фильтр, где обеспечивается фильтрация расплава, поступающего в фильеру. Внутри фильеры расплавы, подаваемые экструдерами, идут по независимым коллекторам, где сливаются в один общий поток на небольшом расстоянии перед выходной щелью. Общий поток расплава выходит из головки через плоскую выходную щель. Зазор щели регулируется 64 термоболтами. Каждый термоболт оснащен нагревательным элементом и температурным датчиком.

Формование пленки-заготовки осуществляется на литьевой машине, состоящей из поливного барабана, водяной ванны, воздушного ножа для прижима пленки к поливному барабану, системы охлаждения воды контура поливного барабана и контура ванны, системы воздушных ножей и валов для осушки плёнки, воздушного ножа для осушки поливного барабана. На поливном барабане осуществляется охлаждение непрерывной ленты полимерного расплава и его кристаллизация. Для получения однородного кристаллического материала необходимо, чтобы расплав постоянно охлаждался в одинаковых условиях по всей ширине барабана. Воздушным ножом осуществляется осушка пленки.

После комплекса поливного барабана перед машиной продольной ориентации (МДО) установлен первый промежуточный намотчик, который используется при пуске и отладке оборудования. В машине продольной ориентации (МДО), осуществляется процесс вытяжки (ориентации) пленки в

продольном направлении. В состав машины МДО (рисунок 3.) входит система валов с приводами и контрольно-измерительные приборы.

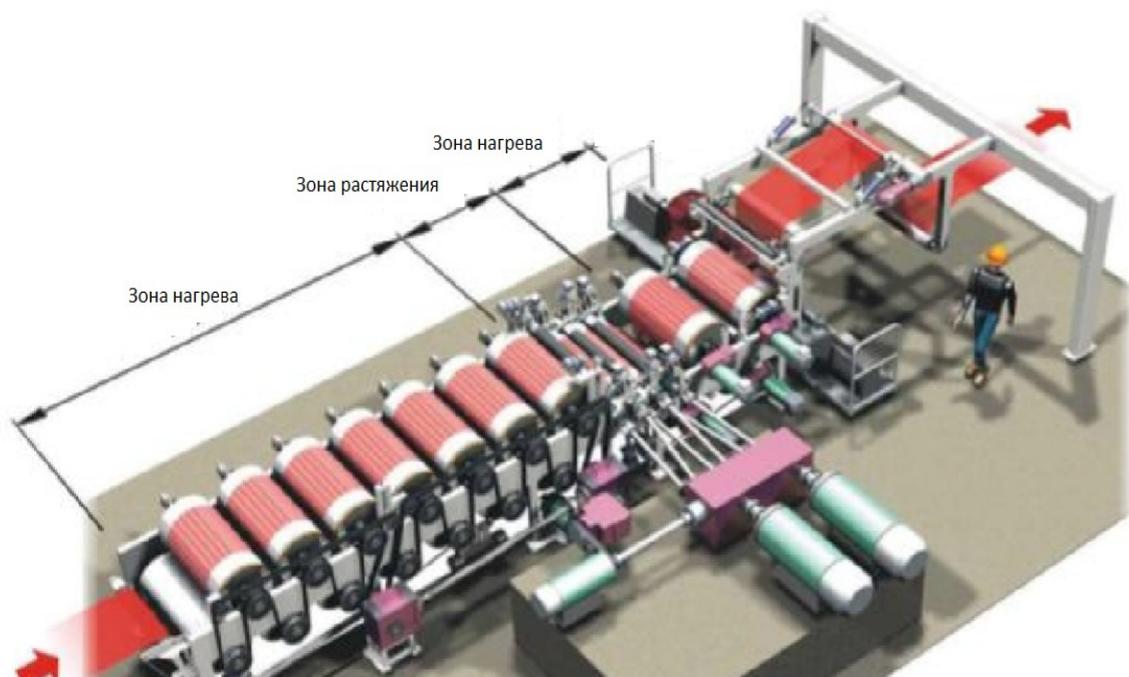


Рисунок 3 – Машина продольной ориентации

Продольно ориентированная пленка через направляющие валки и заправочное устройство подается в установку поперечной ориентации ТДО (рисунок 4.), где края пленки захватываются специальными зажимами (крупками) шириной 120 мм, расположенными на двух автоматически смазываемых замкнутых цепях (шаг цепи 60 мм), движущихся по направляющим рельсам. Пленка движется через зону предварительного нагрева, где производится подогрев пленки горячим воздухом. Через зону растяжения, где пленка растягивается в поперечном направлении. Через зону термофиксации, где снимается внутреннее напряжение в пленке и достигается заданный внешний вид пленки.

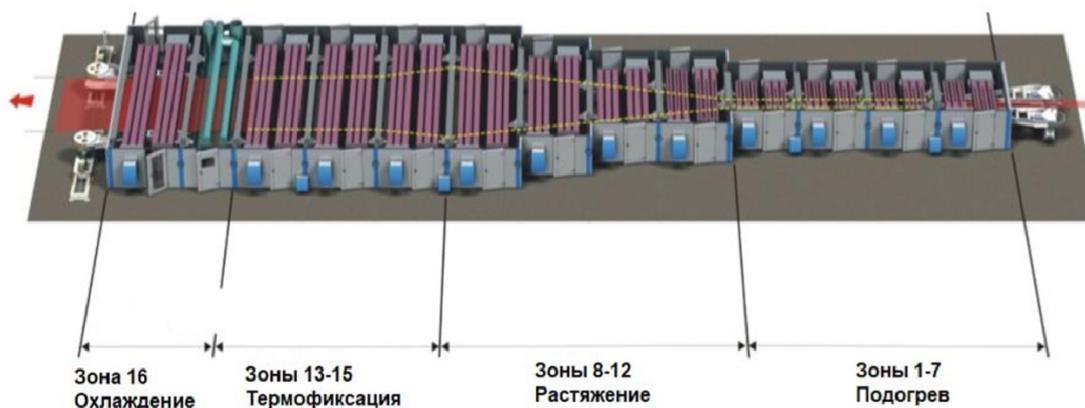


Рисунок 4 – Установка поперечной ориентации

С выхода из установки ТДО двуосноориентированная пленка поступает в систему стана тянущих валов (ПРС), в которую входят все узлы и устройства от выхода ТДО до намотчика. Все валы системы ПРС подразделяются на группы, приводятся во вращение от приводов постоянного тока и выполняют следующие функции:

1. группа 1 производит отбор пленки из установки ТДО, выравнивание полотна, охлаждение и проводку пленки через толщиномер. Включает разравнивающие ролики, охлаждающие валы с внутренним водяным охлаждением;
2. группа 2 обеспечивает удаление отрезанных утолщенных краев (кромки). Включает разравнивающие ролики, тянущий вал;
3. группа 3 обеспечивает пламенную обработку внешней стороны пленки и проводку пленки к следующей группе. Включает охлаждаемый вал и прижимной вал;
4. группа 4 обеспечивает коронную обработку внешней стороны пленки и проводку пленки к следующей группе. Включает направляющий вал охлаждаемый диэлектрический вал и прижимной вал;
5. группа 5 обеспечивает коронную обработку внутренней стороны пленки. Включает охлаждаемый диэлектрический вал и прижимной вал;

6. прижимные валы препятствуют проскальзыванию пленки и удалению воздуха между плёнкой и валом;

7. группа 6 обеспечивает охлаждение и проводку пленки до намотчика. Включает охлаждающий вал с внутренним водяным охлаждением и направляющие валы.

Пленке придаются требуемые заказчиком свойства с помощью станции газопламенной и коронной обработки.

После обработки плёнка наматывается на гильзы. Мостовой кран разгружает их. Образцы пленки поступают в лабораторию и происходит проверка качества. Затем, готовая пленка отправляется на склад.

В данной магистерской работе исследуются двигатели ТДО, некорректная работа которых приводит к деформациям пленки.

2 Оборудование ТДО

2.1 Контроллерное оборудование ТДО

На технологической линии производства БОПП-пленки сбор и передача данных осуществляется тремя ведущими ПЛК. Каждый ПЛК управляет своей частью технологической линии. На исследуемой установке поперечного растяжения (ТДО) используется Siemens S7-1500 (рисунок 5.) [14], характеристики которого представлены в таблице 1.

Таблица 1 – Характеристики контроллера Siemens S7-1500

Рабочая температура	Интерфейсы	Протокол передачи данных	Дискретный ввод/вывод	Аналоговый ввод/вывод	Время выполнения операций	Наработка на отказ	Цена
(минус 40 – 70) °С	Ethernet, RS232, RS485, USB	PROFINET IO, PROFIBUS DP, AS-Interface, OPC UA	10/6	8/4	0,1 мкс	350 000 ч	от 24 458р



Рисунок 5 – Siemens S7-1500

Siemens S7-1500 – необходим для построения различного рода систем автоматизации (низкой и средней степени сложности). Контроллер имеет модульную конструкцию и включает в себя следующие модули:

1. модуль центрального процессора;
2. блоки питания;
3. сигнальные модули;
4. коммуникационные модули;
5. функциональные модули;
6. интерфейсные модули.

Конструкция контроллера выделяется высоким удобством обслуживания:

1. Каждый модуль устанавливается на профильную шину S7-1500 и фиксируются в рабочих положениях винтами. С помощью шинных соединителей осуществляется объединение модулей в одну систему;

2. модули располагаются произвольно в монтажных стойках. Из-за наличия съемных фронтальных соединителей.

Центральный процессор Siemens S7-1500 обладает следующими показателями:

1. Быстродействие (одно из самых высоких на рынке контроллеров);
2. большое количество активных коммуникационных соединений;
3. CPU 15xxC, CPU 15xT-2 DP имеют набор встроенных аналоговых входов/выходов. То есть центральный процессор можно использовать, как готовый блок управления.

Программирование Siemens S7-1500 реализуется по стандарту МЭК 61131-3 в среде TIA Portal.

2.2 Датчики ТДО

Для контроля обрыва пленки используются ультразвуковые датчики обрыва – UC-F77 (рисунок 6.) [15].



Рисунок 6 – Ультразвуковой датчик обрыва

Датчики позволяют точно обнаруживать сложные материалы, такие как прозрачная пленка. Способны регулировать ширину звукового луча и игнорировать нежелательные отражения. Имеют высокую помехоустойчивость.

2.3 Двигатели ТДО

Для прохождения пленки через ТДО используются два асинхронных двигателя на входе установки и два двигателя на выходе – SIEMENS SIMOTICS M-1PH8 (рисунок 7.) [16].

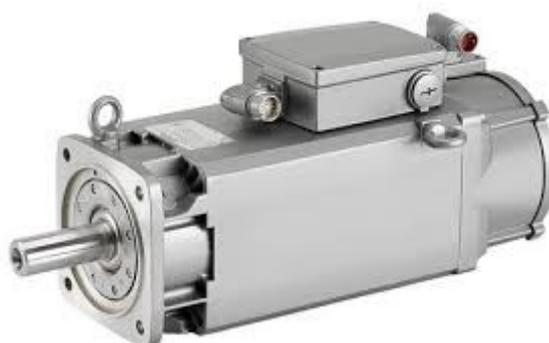


Рисунок 7 – Асинхронный двигатель M-1PH8

Синхронизация двигателей осуществляется с помощью энкодеров и частотных преобразователей Siemens – SINAMICS S120.

2.4 Ректифер ТДО

Siemens Powerblock – преобразовывает переменный ток в постоянный для управления двигателями (рисунок 8.) [17].



Рисунок 8 – Ректифер

2.5 Control Unit ТДО

Управление двигателями производится приводной системой SINAMICS S120 (рисунок 9.) [18]. Управляющий модуль рассчитывает приводы и сетевое питание/рекуперацию. Силовые части (силовые модули, модули питания и модули двигателей) обеспечивают оптимальное преобразование энергии для и двигателя.



Рисунок 9 – SINAMICS S120

На ТДО установлены «Master» и «Slave» SINAMICS S120. «Slave» используются для регулирования скорости и момента двигателей, кроме того, обрабатывают информацию о настройках, включении и отключения двигателя. «Master» синхронизирует работу «Slave» устройств.

3 Схема информационных потоков ТДО

Схема информационных потоков представлена в приложении А.

Данная схема состоит из четырех уровней:

1. Уровень датчиков и исполнительных механизмов. Интерфейсы уровня – 1-WIRE, 4-20Ma, DRIVE-CLIQ. Устройства – датчики (фиксирующие разрыв пленки) и исполнительные механизмы (двигатели). Сигналы – аналоговые и дискретные.

2. Контроллерный уровень. Интерфейс – RS485. Протоколы – PROFIBUS, PROFINET. Устройства – ПЛК (Siemens S7-1500) и модули ввода/вывода. Три ПЛК собирают информацию по всей технологической линии с помощью модулей ввода/вывода с датчиков, в зависимости от неё формируют управляющие сигналы на исполнительные механизмы. Кроме того, передают данные на диспетчерский уровень.

3. Диспетчерский уровень. Интерфейс – RJ-45. Протоколы – Profinet, Ethernet, Modbus TCP/IP. Устройства – персональные компьютеры, сетевые коммутаторы, локальные HMI-панели. Информация с ПЛК поступает в SCADA-пакет (Vijeo Citect), установленный на АРМ оператора. SCADA-пакет позволяет визуализировать работу технологической линии, архивировать её данные, а также управлять ей.

4. СУУТП уровень. На данном уровне производится архивация временных рядов базой данных, построение прогнозов, детектирование аномалий и онлайн мониторинга процесса.

4 Разработка предиктивной модели

4.1 Основные шаги в разработке предиктивной модели

1. Постановка задачи. Определение того, как будет использоваться предиктивная модель, для чего она нужна, её составные части и насколько эффективна интеграция модели в производство.

2. Сбор информации. Чем больше накоплено исторических данных, тем больше моделей для построения прогноза можно применить и тем более эффективной они могут быть.

3. Предварительный анализ. Построение графиков исследуемого процесса для определения каких-либо закономерностей, тенденций или выбросов. Расчет корреляции между данными для определения насколько сильны отношения между ними.

4. Построение прогнозной модели. Для предиктивного обслуживания необходимо выполнять прогнозы на часы/дни/недели вперед.

5. Построение модели для детектирования аномалий. После получения прогноза необходимо определить, насколько будущий процесс стабилен.

6. Выбор моделей. Сравнение моделей в зависимости от метрик качества, от скорости работы и возможности интеграции в производство.

7. Интеграция в технологический процесс. Определение пути потока информации для организации хранения данных, обучения моделей на этих данных и представление информации в удобной для диспетчеров/операторов форме.

4.2 Сбор данных для предиктивной модели

Для реализации предиктивной модели необходимо иметь набор данных исследуемых объектов технологического процесса, проанализировать

его и приступить к разработке. Так как объект исследования – устройство поперечного растяжения, то информация собирается ПЛК ТДО и отправляется в SCADA-пакет Vijeo Citect.

SCADA-пакет Vijeo Citect – программное обеспечение компании Schneider Electric для мониторинга и управления. Резервирует все компоненты системы. Имеет распределенную архитектуру “клиент-сервер”. Хранит трехмесячные данные технологической линии, которые и понадобятся для исследования.

Данные хранятся в формате .hst. Но для проведения анализа необходим табличный формат .csv. Поэтому с помощью программы Citect Trend Reader v1.2.8 производится преобразование формата.

4.3 Постановка задачи

Временной ряд – это множество наблюдений, получаемых последовательно во времени, обычно через равные промежутки [19].

Информация со SCADA представлена многомерным (или одномерным) временным рядом $X = \{x_1, x_2, \dots, x_T\}$, $x_t \in R^m$, где каждый элемент m -мерный вектор, представляющий собой показания m датчиков и исполнительных механизмов в момент времени t . Необходимо из точки t сделать прогноз $\hat{x}_\tau(t)$ величины $x_{t+\tau}$. В каждый момент времени прогноза определить является ли значения с датчиков аномальным или нет.

Исследуемые объекты – двигатели на входе ТДО и на выходе. Данные скорости, температуры, нагрузки по каждому из объектов.

4.4 Требования к временному ряду

Для совершения оперативных прогнозов с помощью авторегрессионных моделей временной ряд должен быть стационарным. По стационарному ряду можно строить прогноз, так как его будущие статистические характеристики не будут отличаться от наблюдаемых

текущих. Под стационарным процессом ξ_t понимается такой, для которого выполняются следующие условия:

$$M[\xi_t] = \text{const} = a, \quad (1)$$

$$D[\xi_t] = \sigma^2, \quad (2)$$

$$r(t_i, t_j) = r(\tau), \quad (3)$$

где $M[\xi_t]$ – математическое ожидание;

$D[\xi_t]$ – дисперсия;

$r(t_i, t_j)$ – автокорреляционная функция;

$\tau = t_i - t_j$.

Величина автокорреляционной функции зависит от величины сдвига τ , значение a определяет постоянный уровень, относительно которого варьируют значения временного ряда, а постоянная величина σ^2 характеризует размах этой вариации.

Для проверки на стационарность используется тест единичного корня Дики-Фуллера. Делается предположение о виде процесса, породившего данный временной ряд. Строится вспомогательная модель и проверяются гипотезы о коэффициентах этой модели. После чего вывод о стационарности/нестационарности исходного ряда.

Выдвигаются две гипотезы: нулевая гипотеза (H_0) – имеется единичный корень, то есть ряд нестационарен; альтернативная гипотеза (H_1) – нулевая гипотеза отвергается, временной ряд не имеет единичного корня, значит он стационарен. Вывод интерпретируется, используя p -value. Если p -value больше 0.05, то нулевая гипотеза не отклоняется и ряд нестационарный, если p -value меньше или равен 0.05, то отвергается нулевая гипотеза из чего следует, что ряд стационарный.

На рисунках 10, 11, 12, 13 временные ряды скорости двигателя, тока, температуры и момента проверяются на стационарность, автокорреляцию и частичную автокорреляцию.

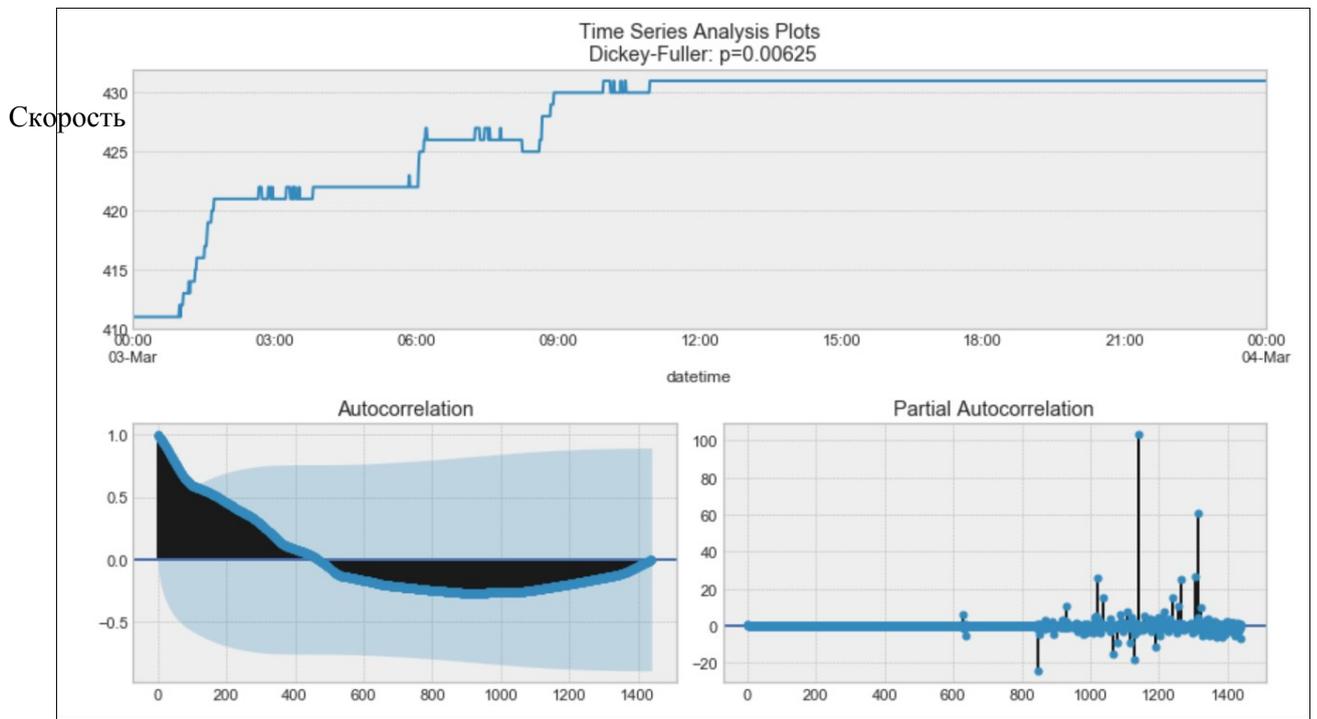


Рисунок 10 – Тест на стационарность ряда (SD)

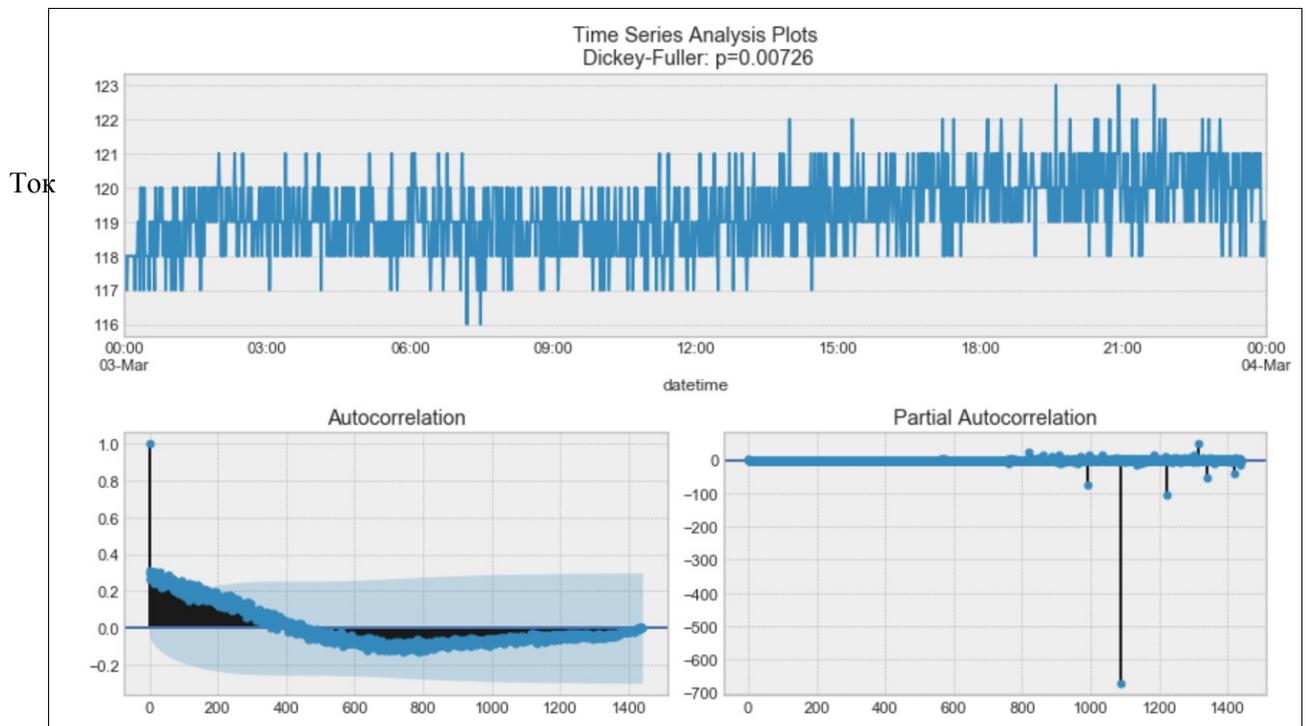


Рисунок 11 – Тест на стационарность ряда (CU)

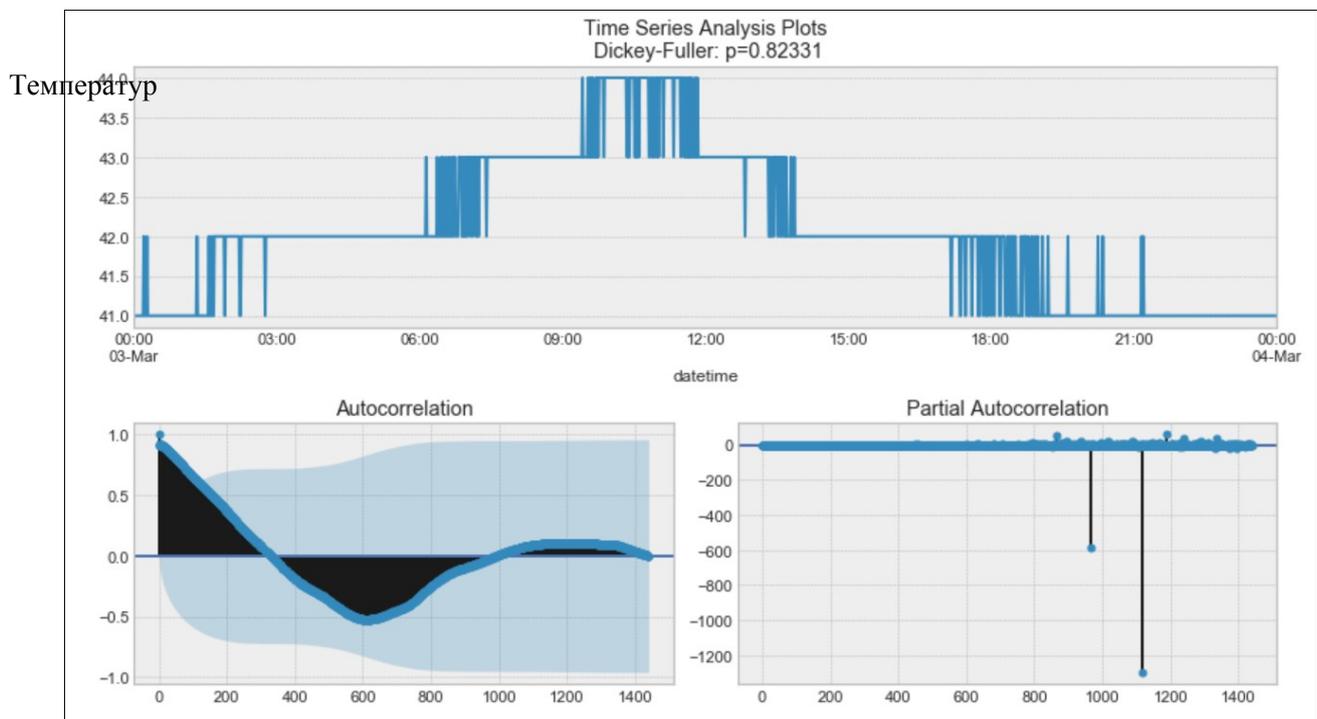


Рисунок 12 – Тест на стационарность ряда (TE)

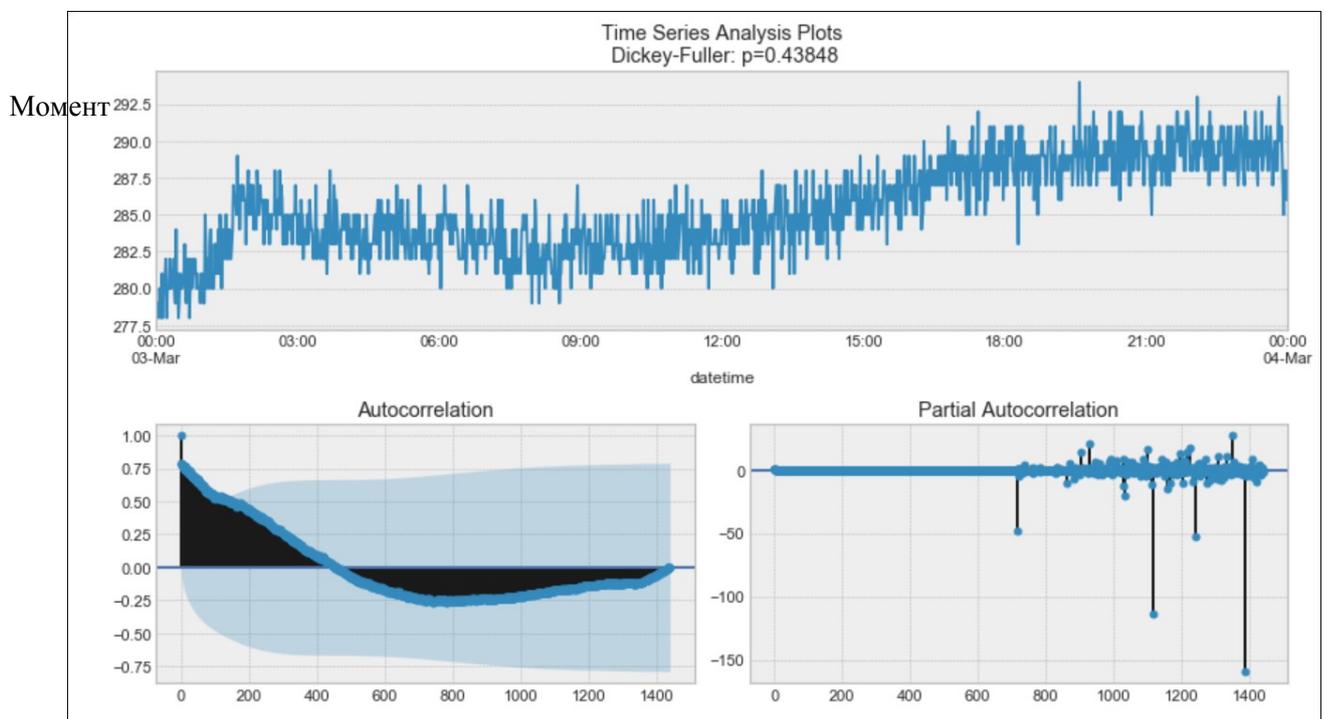


Рисунок 13 – Тест на стационарность ряда (NM)

Временные ряды температуры и момента двигателя не проходят тест на стационарность. Соответственно необходимо провести операцию дифференцирования, которая подразумевает вычитание ряда из самого себя с лагом в один шаг:

$$y_1, \dots, y_T \rightarrow \dot{y}_2, \dots, \dot{y}_T, \quad (4)$$

$$\dot{y}_2 = y_t - y_{t-1}, \quad (5)$$

Тесты на стационарность после преобразования приведены на рисунках 14 и 15:

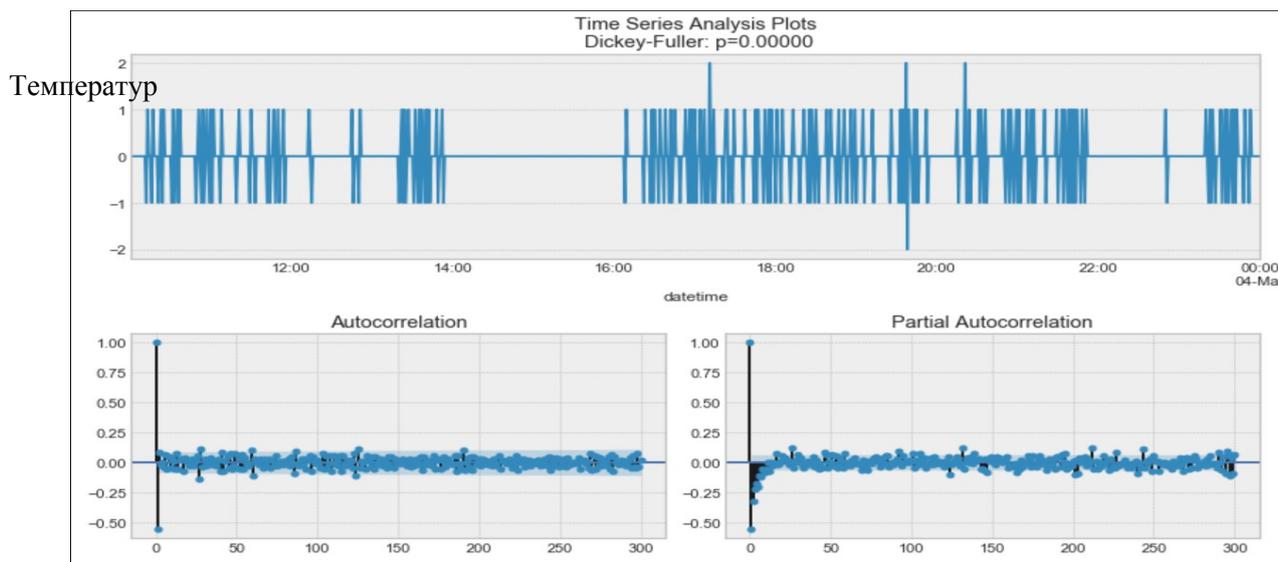


Рисунок 14 – Тест на стационарность ряда ТЕ после преобразований

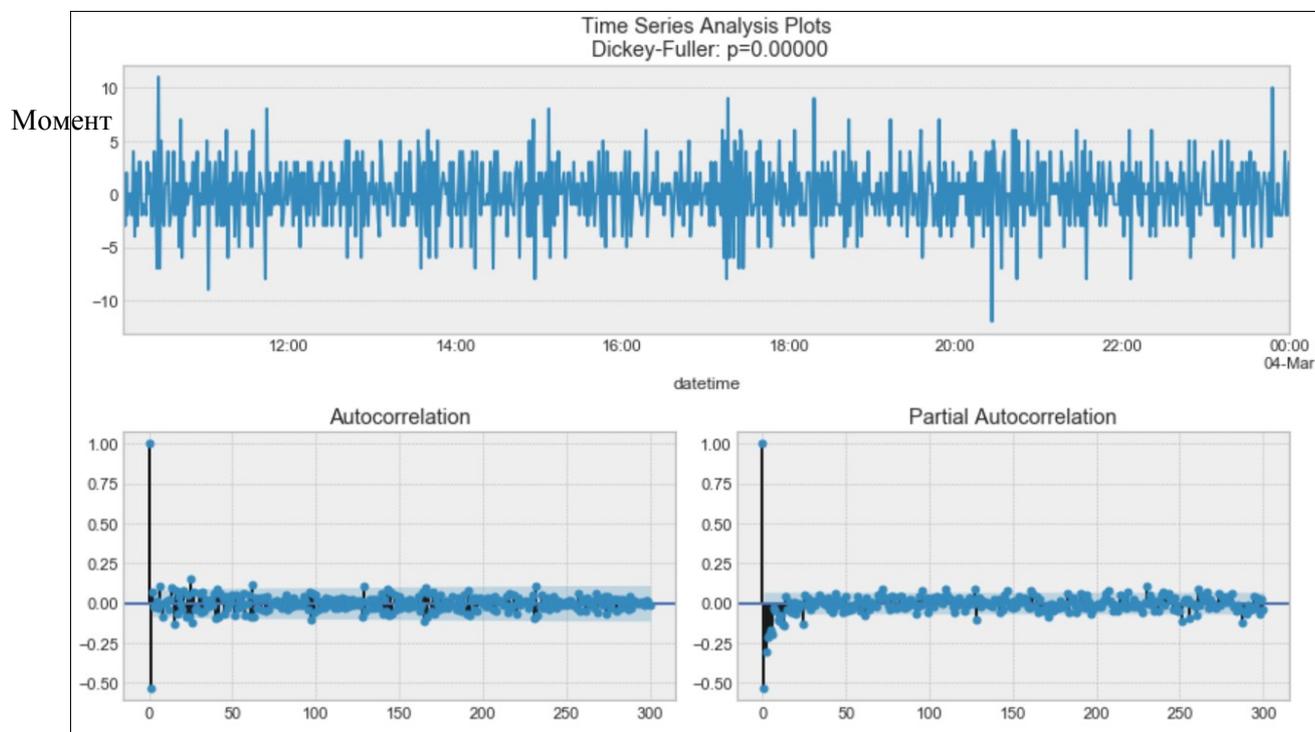


Рисунок 15 – Тест на стационарность ряда NM после преобразований

4.5 Стек технологий для исследования

Для анализа временных рядов существует несколько классических высокоуровневых языков программирования – Matlab, R, Python, Julia. В

случаях, когда решение задачи предполагает статистическую обработку данных, работу с графиков и т.п., эти языки применяются чаще других.

В данном исследовании используется язык Python, веб-оболочка для IPython – Jupyter Notebook, позволяющий объединить код, текст и изображения. Анализ осуществляется следующими библиотеками:

1. NumPy – модуль для сложных вычислений. Предоставляет функционал сравнимый с Matlab – базовые методы для манипуляции с большими массивами и матрицами [20].

2. Pandas – библиотека для обработки и анализа данных. Строится поверх библиотеки NumPy. Предоставляет специальные структуры данных и операции для манипулирования числовыми таблицами и временными рядами – Series, DataFrame [21].

3. Matplotlib – библиотека для визуализации данных, как двумерной графикой, так и трехмерной [22].

4. Seaborn – библиотека для визуализации данных основанная на matplotlib. Позволяет использовать больше видов графиков относительно matplotlib [23].

5. SciKit-learn – библиотека предоставляет реализацию ряда алгоритмов машинного обучения [24].

6. Datetime – библиотека используется для работы с датами и временем [25]

7. Statsmodels – библиотека, предоставляющая классы и функции для оценки различных статистических моделей, а также для проведения статистических тестов и исследования статистических данных [26].

8. Facebook prophet – библиотека, разработанная для прогнозирования большого числа различных бизнес-показателей. Не требует от аналитиков глубоких знаний устройства предсказательных моделей [27].

9. Amazon GluonTS – библиотека, разработанная Amazon для анализа временных рядов. Имеются инструменты, необходимые для построения и обучения наиболее распространенных архитектур нейронных

сетей, механизмы для загрузки и предварительной обработки данных, готовые прогнозные модели, и инструменты для оценки и сравнения различных моделей [28].

10. Tsfresh – библиотека для генерирования признаков. Например, максимальные, минимальные значения, p-value, коэффициенты наклона линии тренда и т.д. [29].

11. Pmdarima – библиотека, включающая в себя эквивалент функции auto.arima из R, коллекцию статистических тестов стационарности и сезонности, различные операции – дифференцирование и обратное дифференцирование временных рядов, встроенные наборы временных рядов и много другое [30].

12. Hmmlearn – библиотека для использования скрытых марковских моделей [31].

13. CatBoost – библиотека реализации градиентного бустинга на деревьях решений с открытым исходным кодом от компании Яндекс. Алгоритм, применяемый для ранжирования и прогнозирования, а также рекомендательных технологий [32].

14. XGBoost – библиотека для прогнозирования, классификации, в основе которого лежит алгоритм градиентного бустинга деревьев решений [33].

15. Keras – библиотека машинного обучения. Используется для решения задач компьютерного зрения, обработка естественного языка, прогнозирования временных рядов и т.д. [34].

4.6 Метрики качества

Основными характеристиками качества статистических моделей, выбранных для прогнозирования, являются показатели точности. Понятие точности характеризует степень близости смоделированных (расчетных по определенной модели прогноза) значений \hat{y} в их совокупности к исходным

фактическим данным y . Чем меньше отличия теоретических значений от эмпирических данных, тем выше качество модели [35].

Таблица 2 – Наиболее используемые метрики качества

Коэффициент детерминации	$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \sum_{i=1}^n y_i)^2}$
Среднее абсолютное отклонение	$MAE = \frac{\sum_{i=1}^n y_i - \hat{y}_i }{n}$
Среднеквадратическая ошибка	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
Средняя абсолютная процентная ошибка	$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{ y_i - \hat{y}_i }{y_i}$

4.7 Подготовка данных

Значения скорости, нагрузки, температуры двигателей снимаются с периодичностью в 1 секунду. Количество наблюдений 500 000 строк. Для начала анализа необходимо привести их к подходящей форме. Исключить пропуски в данных, оставить время и измеряемый параметр (рисунок 14.):

До преобразований							После преобразований	
	Time	Raw Value	Scaled Value	Description	State		datetime	Motor speed
0	03.03.2020 00:00:00.000 GMT	26351	411	734375 M/MN	NaN		2020-03-03 00:00:00	411
1	03.03.2020 00:00:01.000 GMT	26351	411	734375 M/MN	NaN		2020-03-03 00:00:01	411
2	03.03.2020 00:00:02.000 GMT	26351	411	734375 M/MN	NaN		2020-03-03 00:00:02	411
3	03.03.2020 00:00:03.000 GMT	26351	411	734375 M/MN	NaN		2020-03-03 00:00:03	411
4	03.03.2020 00:00:04.000 GMT	26351	411	734375 M/MN	NaN		2020-03-03 00:00:03	411

Рисунок 16 – Преобразование данных

Получившийся временной ряд имеет большое количество выбросов и частое изменение тренда. Происходит это из-за того, что производимая пленка имеет разные свойства в зависимости от требования заказчика. Соответственно определенные свойства пленки получаются при определенных скоростях. Изменение вида пленки, а также скорости технологической линии устанавливается оператором раз в сутки из чего

следует, что имеет смысл прогнозировать работу двигателей на несколько часов вперед, а в качестве исторических данных использовать информацию за сутки с частотой в одну минуту (рисунок 11.):

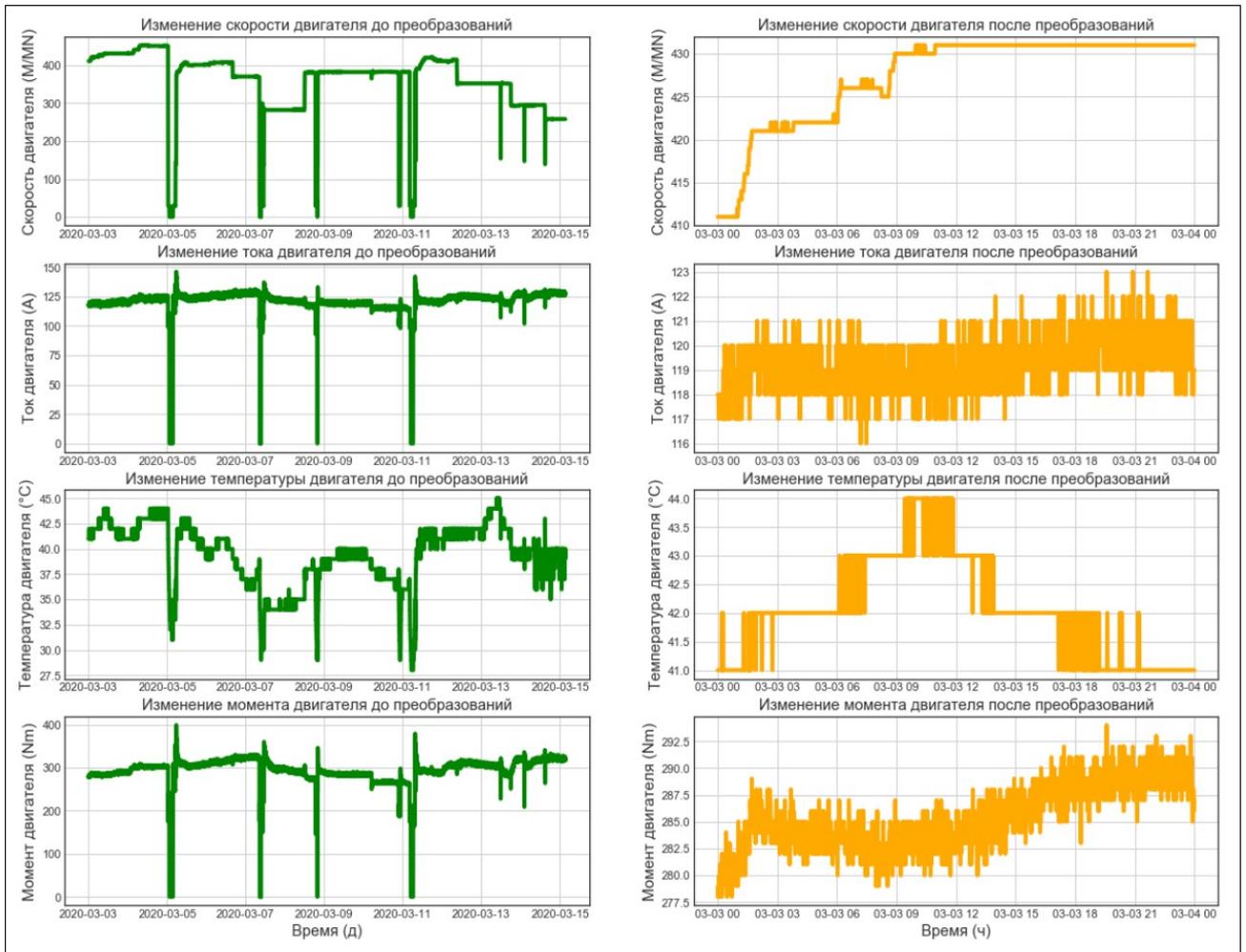


Рисунок 17 – Графики изменений скорости двигателя

4.8 Прогнозные модели

После подготовки данных определяется модель прогнозирования. Модели прогнозирования временных рядов классифицируются на несколько групп (рисунок 11.) [36]:

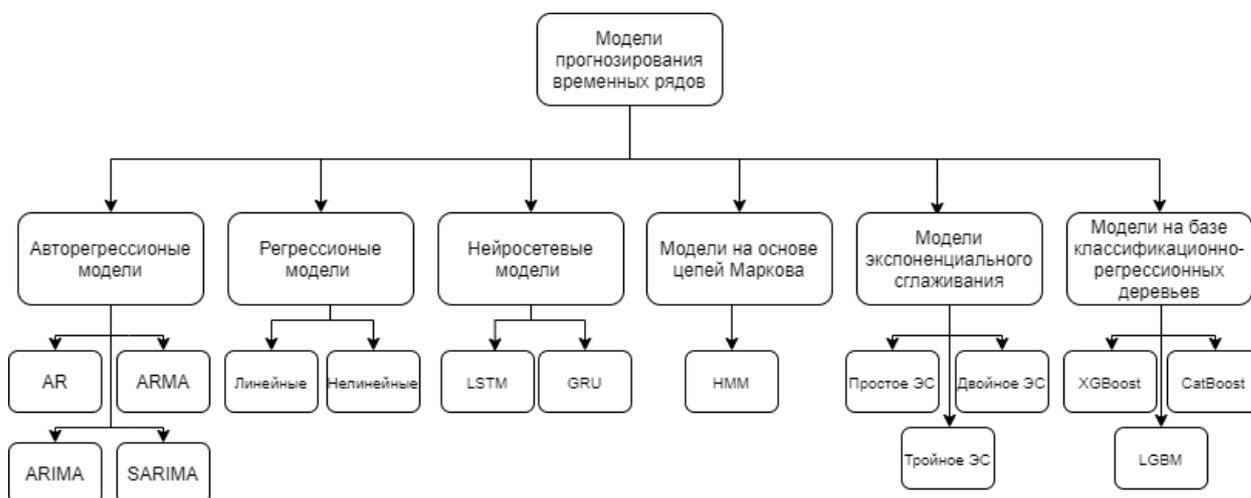


Рисунок 18 – Модели прогнозирования

4.8.1 Методы скользящего среднего (moving average, MA)

Метод заключается в предположении, что будущее значение переменной зависит от среднего n её предыдущих значений:

$$\hat{y}_t = \frac{1}{k} \sum_{n=0}^{k-1} y_{t-n}, \quad (6)$$

где k – размер окна выборки;

\hat{y}_t – предсказанное значение;

y_{t-n} – предыдущее значение ряда.

Таблица 3 – Предсказание на шаг вперед методом MA

	SD, M/MN	CU, A	TE, °C	NM, Nm
MA	427.3470139	119.2956	42.052	285.38168

Одним из недостатков скользящей средней является присвоение при расчете всем наблюдениям одинакового веса. Из чего следует появление модели взвешенной скользящей средней, внутри которой наблюдениям придаются различные веса, в сумме дающие единицу:

$$\hat{y}_t = \frac{1}{k} \sum_{n=0}^{k-1} w_n y_{t-n}, \quad (7)$$

где w_n – вес наблюдения.

Таблица 4 – Предсказание на шаг вперед методом WMA

	SD, M/MN	CU, A	TE, °C	NM, Nm
WMA	431.0	119.0	41.0	287.3

4.8.2 Методы экспоненциального сглаживания

4.8.2.1 Экспоненциальное сглаживание

Временной ряд сглаживается с помощью взвешенной скользящей средней, в которой веса подчиняются экспоненциальному закону, то есть более ранним наблюдениям придаются меньшие веса, чем более поздним.

Экспоненциальное сглаживание предусматривает постоянное обновление модели за счет наиболее свежих данных. В сглаженном виде новый прогноз можно представить как взвешенное среднее последнего наблюдения величины в момент времени t и её прежнего прогноза за этот же период. Вес α присваивается последнему наблюдаемому значению, а вес $(1-\alpha)$ – прогнозу:

$$\hat{y} = \alpha \cdot y_t + (1-\alpha) \cdot \hat{y}_{t-1}$$

где \hat{y} – прогнозируемое значение на период $t + 1$;

y_t – фактическое значение ряда в момент времени t ;

\hat{y}_{t-1} – прежний прогноз на момент времени t ;

вес α – сглаживающий фактор.

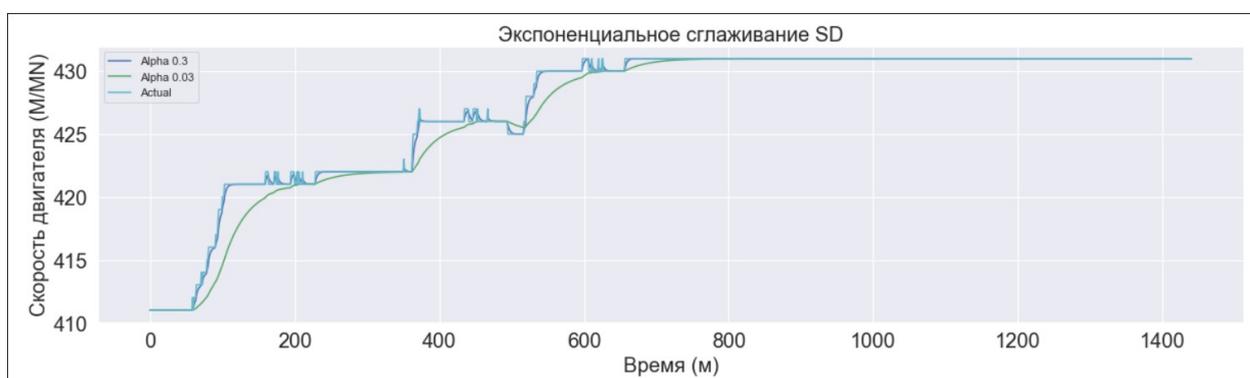


Рисунок 19 – Экспоненциальное сглаживание SD

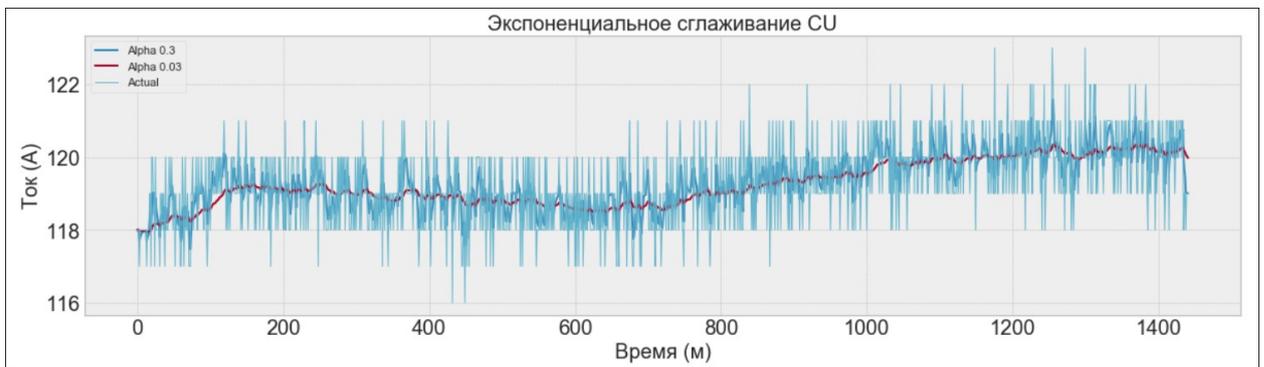


Рисунок 20 – Экспоненциальное сглаживание CU

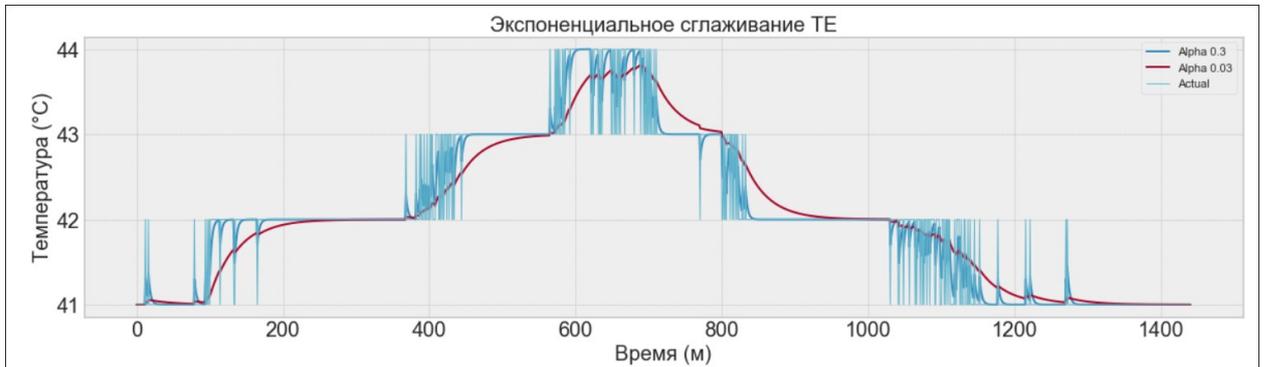


Рисунок 21 – Экспоненциальное сглаживание TE

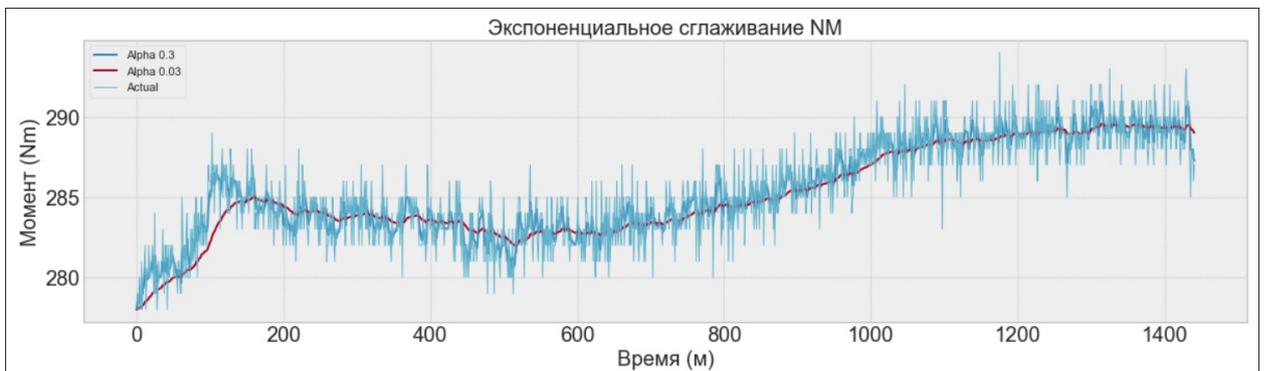


Рисунок 22 – Экспоненциальное сглаживание NM

Таблица 5 – Предсказание на шаг вперед методом ЭС

	SD, M/MN	CU, A	TE, °C	NM, Nm
ЭС	431.0	119.007	41.0	287.2695

Теоретическое обоснование метода представлено в работе [37].

4.8.2.2 Двойное экспоненциальное сглаживание

Методы скользящего среднего, взвешенного скользящего среднего, простого экспоненциального сглаживания позволяют сделать предсказание на шаг вперед. Расширение экспоненциального сглаживания – двойное

экспоненциальное сглаживание. Добавляются две составляющие – уровень l и тренд b :

$$l_x = \alpha y_x + (1 - \alpha)(l_{x-1} + b_{x-1}), \quad (8)$$

$$b_x = \beta (l_x - l_{x-1}) + (1 - \beta)b_{x-1}, \quad (9)$$

$$\hat{y}_{x+1} = l_x + b_x, \quad (10)$$

где α, β – веса в двойном экспоненциальном сглаживании.

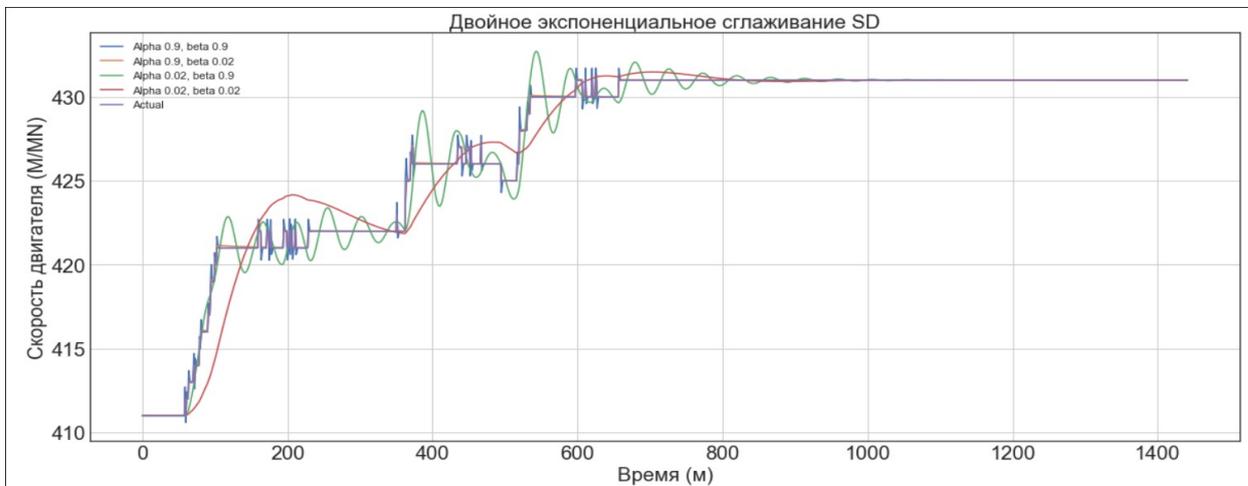


Рисунок 23 – Двойное экспоненциальное сглаживание SD

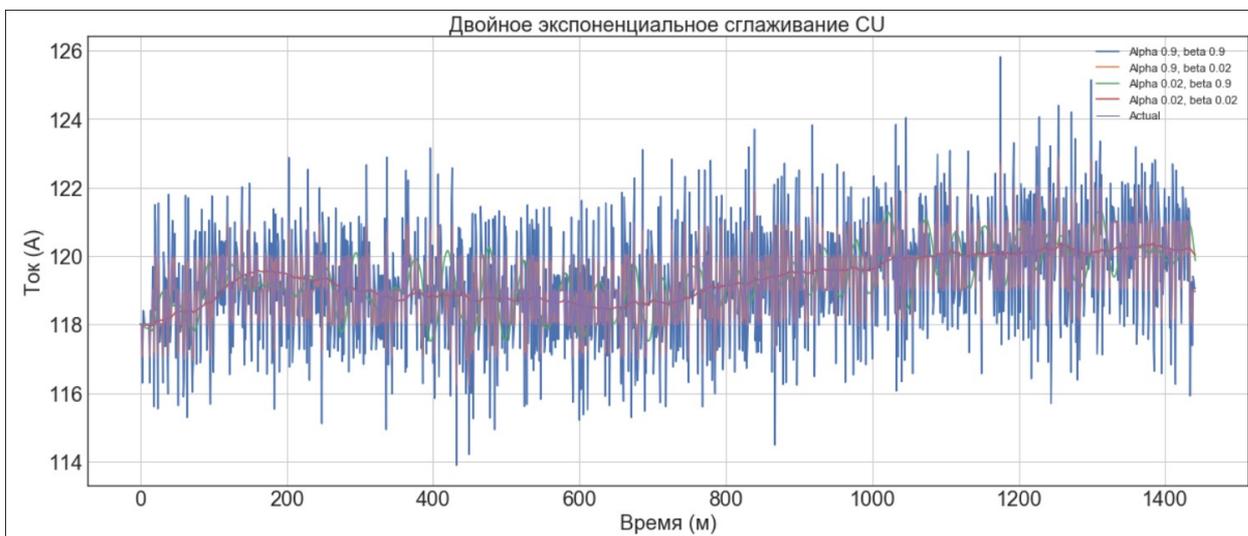


Рисунок 24 – Двойное экспоненциальное сглаживание CU

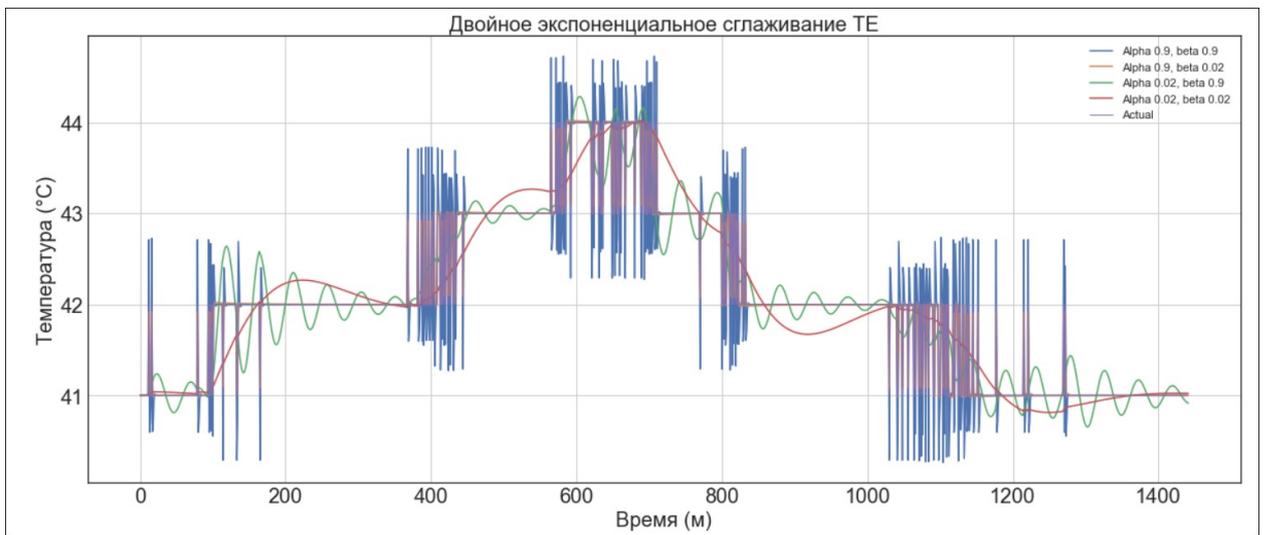


Рисунок 25 – Двойное экспоненциальное сглаживание TE

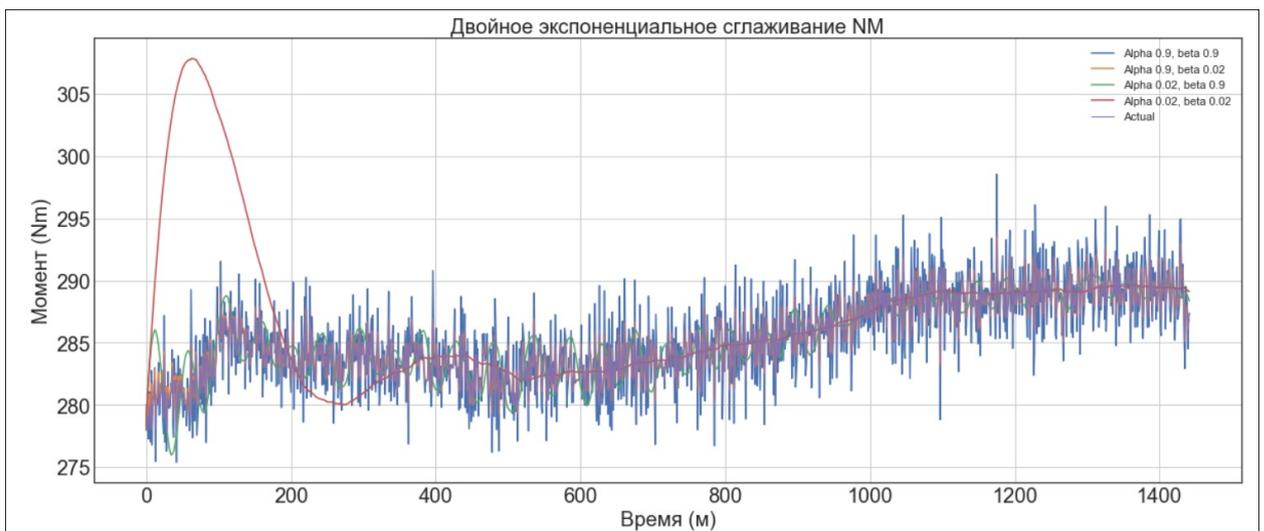


Рисунок 26 – Двойное экспоненциальное сглаживание NM

Таблица 6 – Предсказание на два шага вперед методом ДЭС

	SD, M/MN	CU, A	TE, °C	NM, Nm
ДЭС	431.0	118.98	40.99992	286.87
	431.0	118.95	40.99985	286.82

Теоретическое обоснование метода представлено в работе [37].

4.8.2.3 Тройное экспоненциальное сглаживание (Holt-Winters)

Помимо двух компонентов, добавленных в двойном экспоненциальном сглаживании, в тройном экспоненциальном сглаживании добавляется третья компонента – сезонность. Сезонная компонента объясняет повторяющиеся колебания вокруг уровня и тренда. Она

характеризуется длиной сезона – периодом, после которого начинается повторение колебаний:

$$l_x = \alpha(y_x - s_{x-L}) + (1-\alpha)(l_{x-1} - b_{x-1}), \quad (11)$$

$$b_x = \beta(l_x - l_{x-1}) + (1-\beta)b_{x-1}, \quad (12)$$

$$s_x = \gamma(y_x - l_x) + (1-\gamma)s_{x-L}, \quad (13)$$

$$\hat{y}_{x+1} = l_x + mb_x + s_{x-L+1+(m-1)modL}, \quad (14)$$

где α, β, γ – веса в двойном экспоненциальном сглаживании.

Предсказание производится на 200 шагов вперед. В роли метрики качества используется MAPE. Параметры для модели подбирались с помощью кросс-валидации.

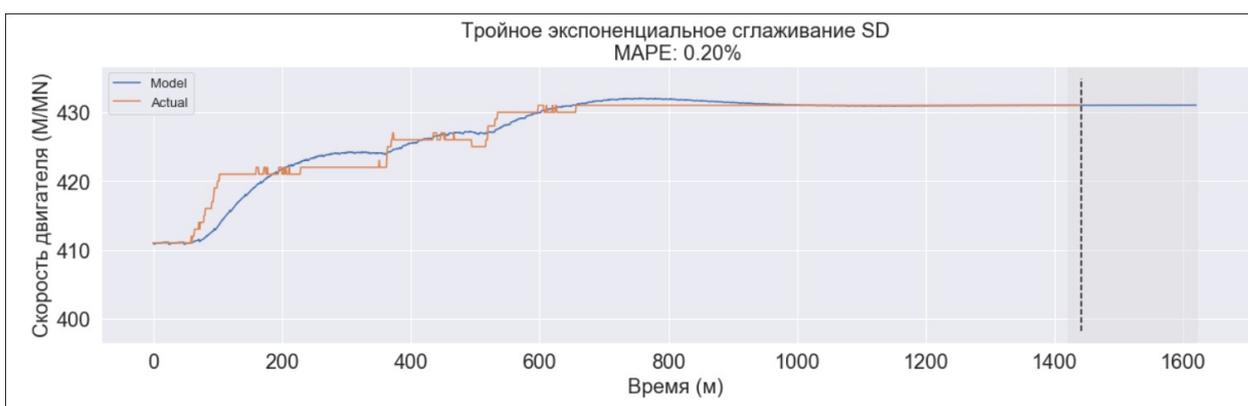


Рисунок 27 – Тройное экспоненциальное сглаживание SD

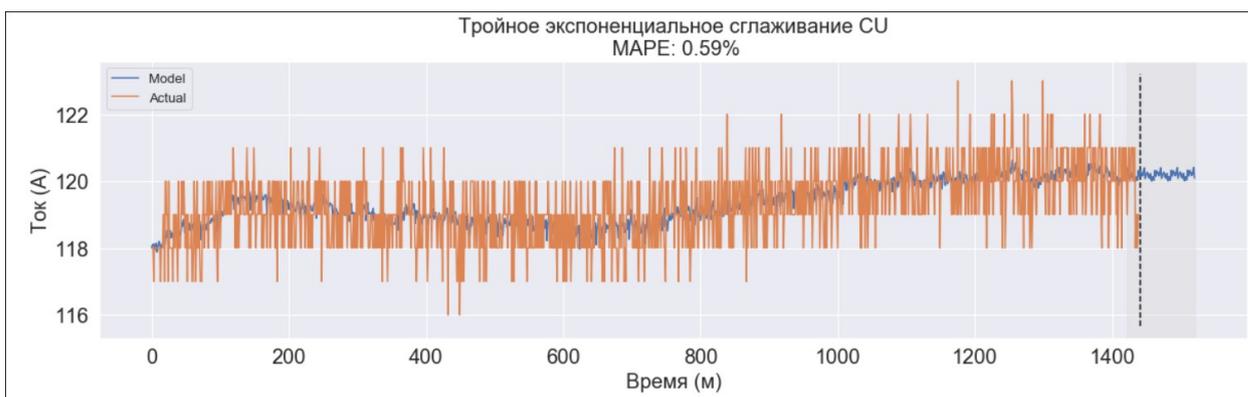


Рисунок 28 – Тройное экспоненциальное сглаживание CU

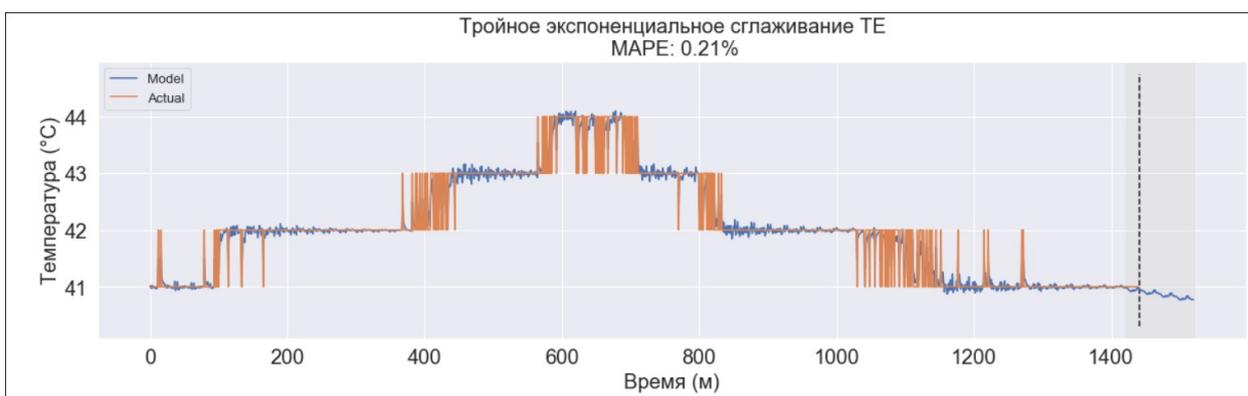


Рисунок 29 – Тройное экспоненциальное сглаживание ТЕ

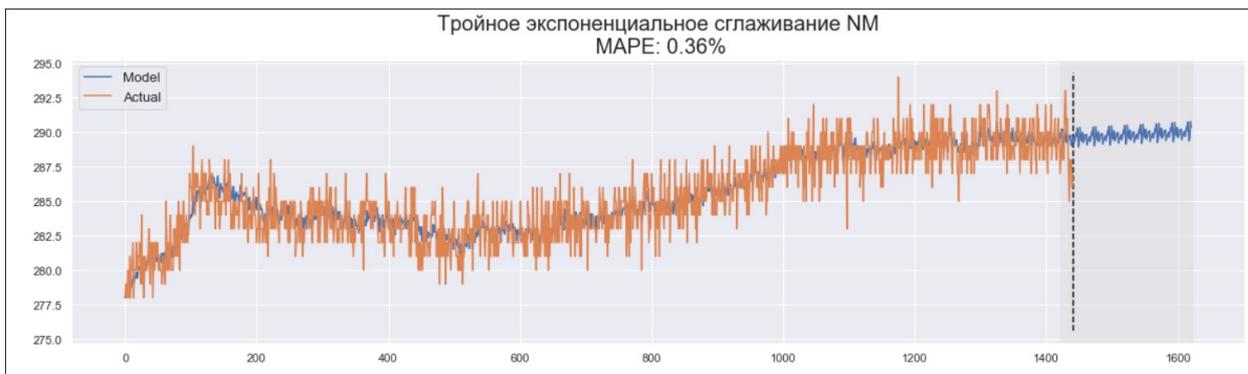


Рисунок 30 – Тройное экспоненциальное сглаживание NM

Теоретическое обоснование метода представлено в статье [38].

4.8.3 Авторегрессионные модели

4.8.3.1 AR (autoregressive) модель

Модель стационарного процесса, выражающая значение показателя y_t в виде линейной комбинации конечного числа предшествующих значений этого показателя и аддитивной случайно составляющей, называются моделью авторегрессии или авторегрессионной моделью:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t, \quad (15)$$

где a_j – коэффициенты модели, p – порядок авторегрессии, ε_t – случайная составляющая, имеющая нормальное распределение с нулевым математическим ожиданием.

На рисунке 31 приведены предсказания с помощью модели авторегрессии:

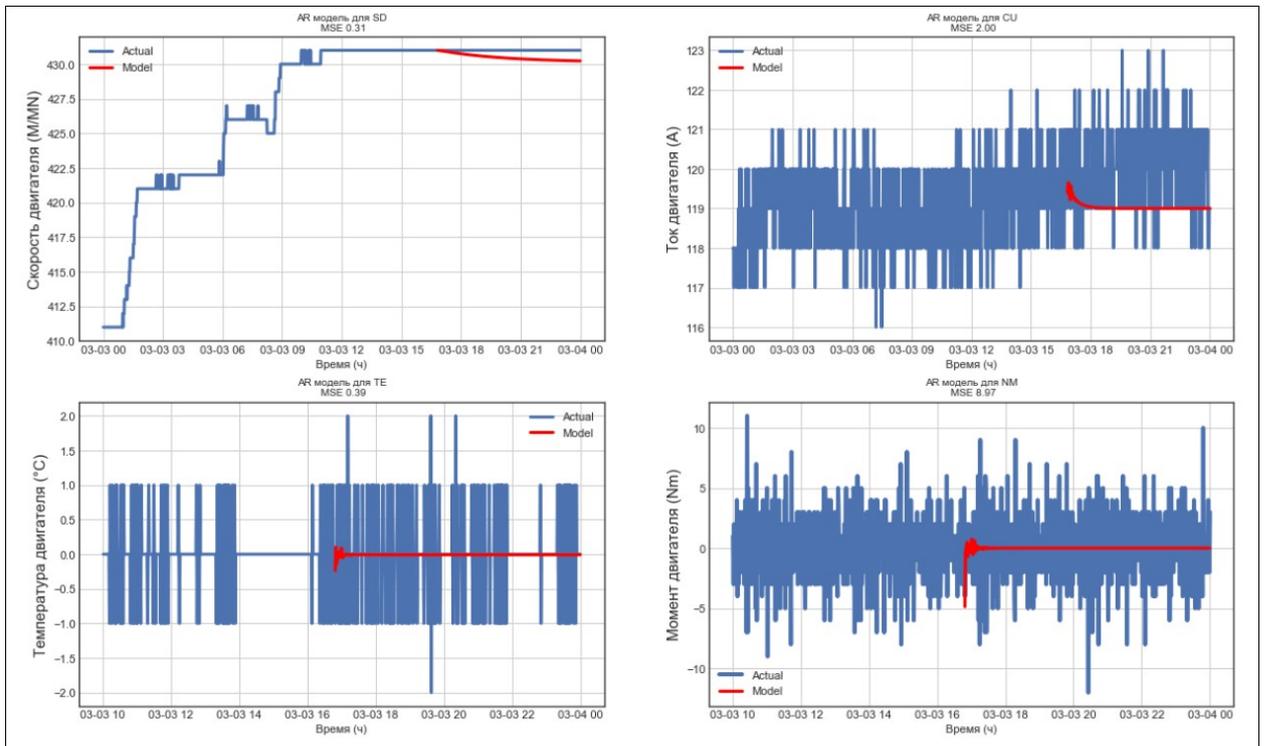


Рисунок 31 – AR модели

Теоретическое обоснование метода представлено в работе [39].

4.8.3.2 ARIMA (autoregressive integrated moving average) модель

Комбинация авторегрессионной модели порядка (см. 4.8.3.1) p и модели скользящего среднего порядка (см. 4.8.1) q дает смешанную авторегрессионную модель скользящего среднего (ARMA(p, q)). ARMA, как и AR модель, используется только для стационарных рядов. Для работы с нестационарными рядами формируется модель ARIMA, тем что в модель ARMA включается оператор взятия простых разностей порядка d . Прогнозирование с помощью ARIMA в явном виде выглядит следующим образом:

$$\hat{y}_{t+l} = \hat{a}_1 y_{t+l-1} + \hat{a}_2 y_{t+l-2} + \dots + \hat{a}_p y_{t+l-p} + \dots + \varepsilon_t - \hat{b}_1 \varepsilon_{t+l-1} - \hat{b}_2 \varepsilon_{t+l-2} - \dots \quad (16)$$

где \hat{a} и \hat{b} – коэффициенты модели, p – порядок авторегрессии, q – порядок скользящего среднего.

Поиск гиперпараметров – p, q, d производится итерационным способом. Сравнивается метрика качества (MAPE) каждого набора гиперпараметров, набор с наименьшей MAPE используется в модели.

На рисунках 32, 33, 34, 35 представлены графики предсказания с помощью модели ARIMA:

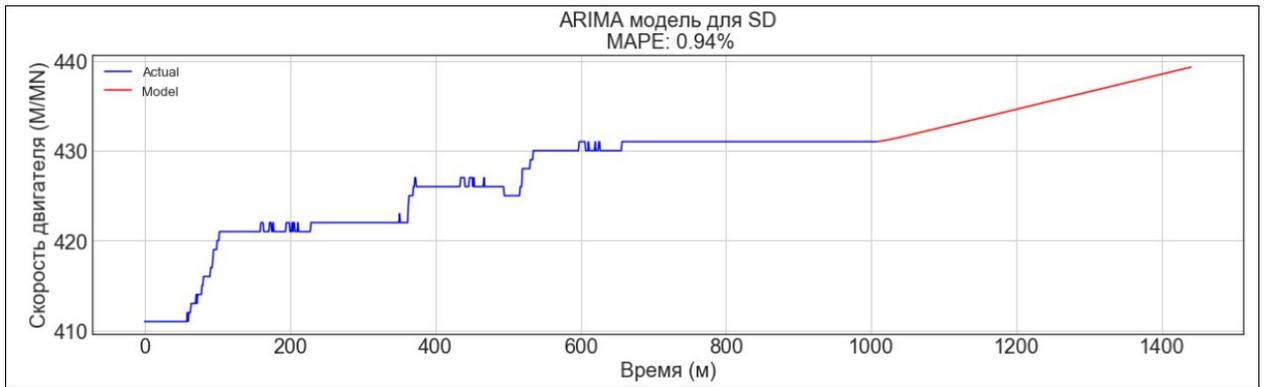


Рисунок 32 – ARIMA модель для скорости двигателя

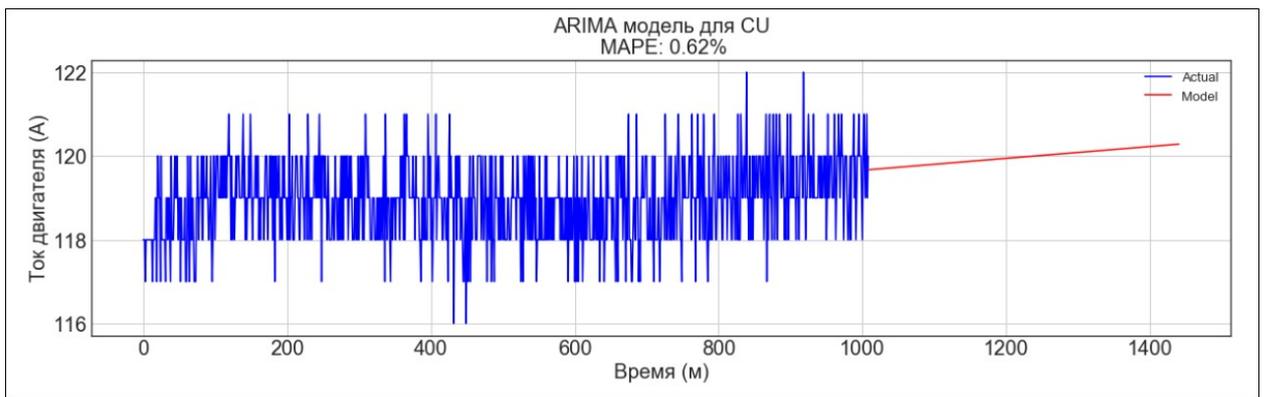


Рисунок 33 – ARIMA модель для тока двигателя

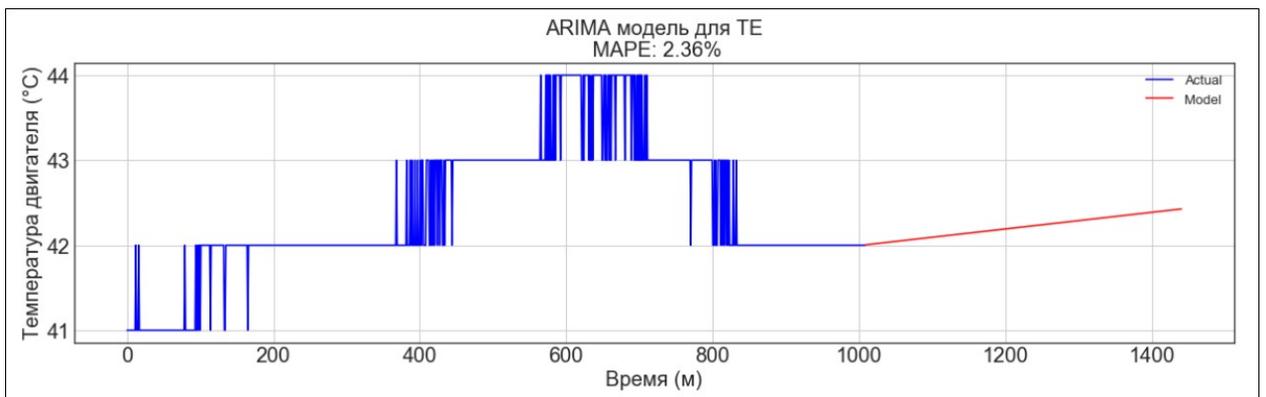


Рисунок 34 – ARIMA модель для температуры двигателя

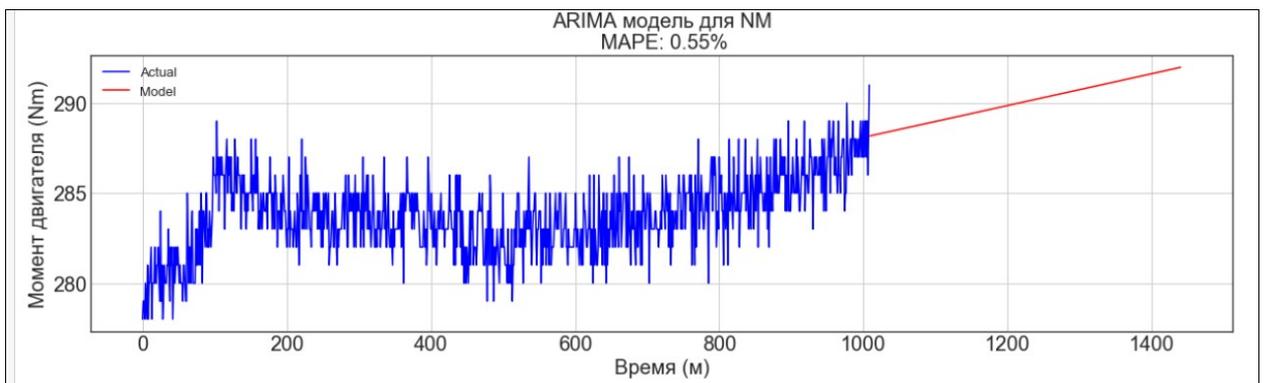


Рисунок 35 – ARIMA модель для момента двигателя

Теоретическое обоснование метода представлено в статье [40].

4.8.3.3 SARIMA (seasonal autoregressive integrated moving average)

модель

Для построения авторегрессионной модели на временных рядах с присутствием сезонного компонента используется расширение модели ARIMA – SARIMA, где S – seasonal. Сезонность – некоторый повторяющийся паттерн, наблюдаемый во временных рядах. В общем виде модель выглядит так:

$$ARIMA(p, d, q)(P, D, Q)_s \quad (17)$$

где p, d, q – параметры ARIMA, $(P, D, Q)_s$ – группа новых параметров SARIMA. P – порядок сезонной составляющей SAR(P), D – порядок интегрирования сезонной составляющей, Q – порядок сезонной составляющей SMA(Q), s – размерность сезонности.

На рисунках 36, 37, 38, 39 представлены графики предсказания с помощью модели SARIMA:

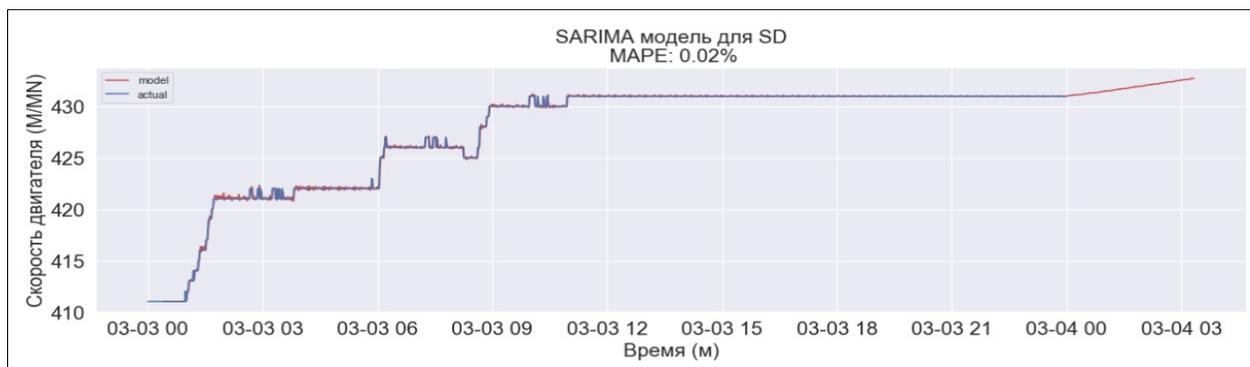


Рисунок 36 – SARIMA модель для скорости двигателя

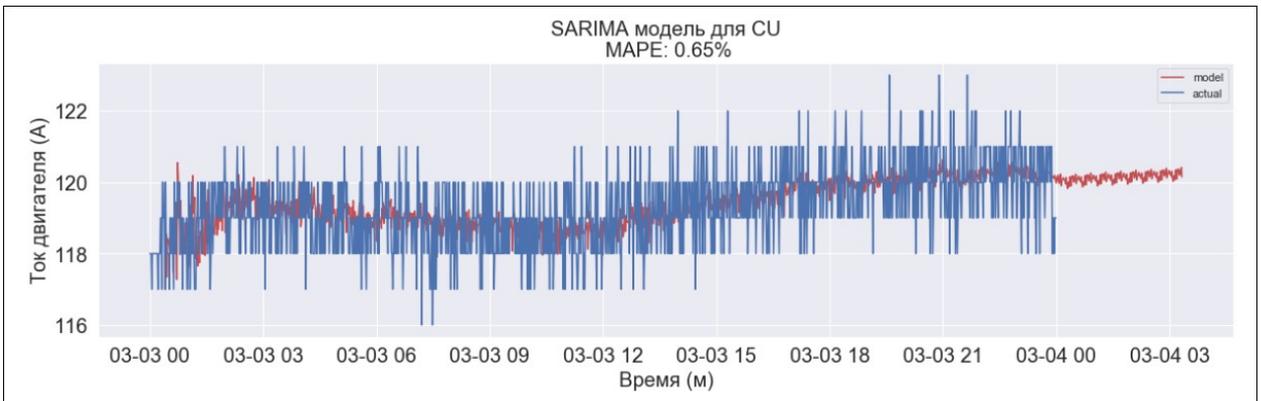


Рисунок 37 – SARIMA модель для тока двигателя

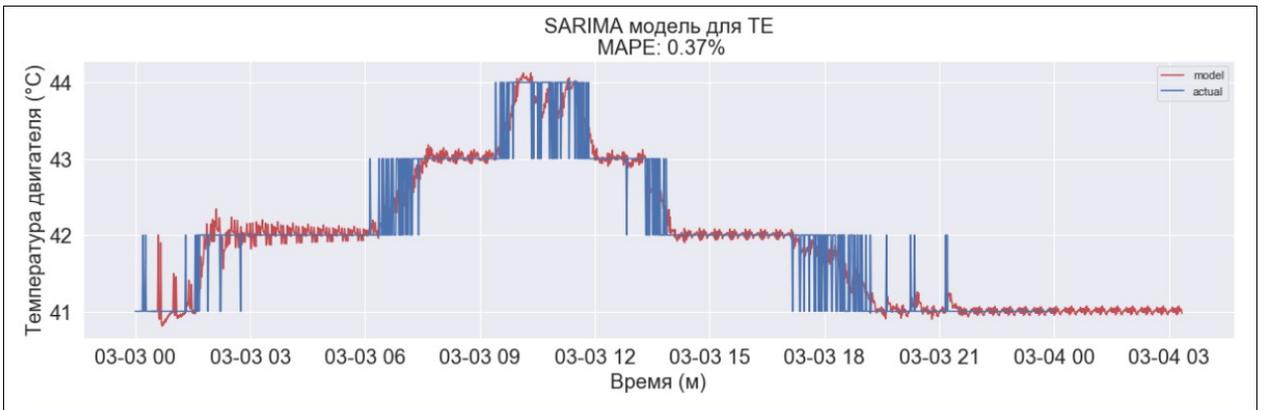


Рисунок 38 – SARIMA модель для температуры двигателя

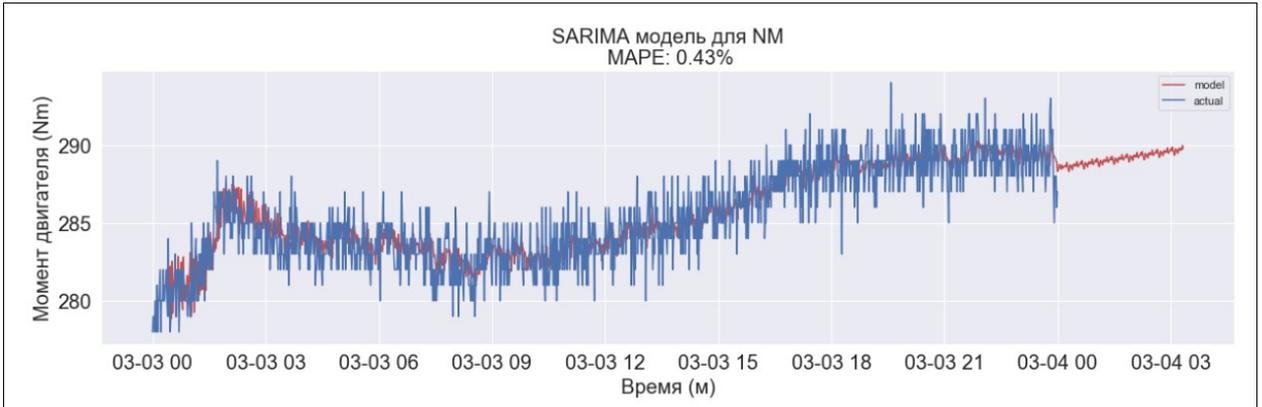


Рисунок 39 – SARIMA модель для момента двигателя

Теоретическое обоснование метода представлено в работе [41].

4.8.4 Скрытая марковская модель (Hidden Markov model)

Скрытая марковская модель – модель процесса, в которой процесс считается марковским, но неизвестно в каком состоянии находится система (состояния скрыты), но каждое состояние может с некоторой вероятностью произвести событие, которое можно наблюдать. Модель состоит из набора

наблюдаемых переменных $X = x_1, \dots, x_n$, где $x_k \in R^d$, и набора латентных (или скрытых) переменных $T = t_1, \dots, t_n$, где $t_k \in [0,1]^K$, $\sum_{i=1}^K t_{ki} = 1$. Скрытые переменные являются бинарными и кодируют K состояний: переменные состояния. Значения наблюдаемого вектора x_k в момент времени k зависят только от скрытого состояния t_k , а t_k зависит только от t_{k-1} .

Применительно к временным рядам скрытое состояние характеризуется периодами, когда математическое ожидание и дисперсия не меняются. Количество скрытых состояний выбирается исходя из наибольшего правдоподобия модели. Правдоподобия представлены в таблице 6. Предсказания на рисунках 40, 41, 42, 43.

Таблица 6 – выбор количества скрытых состояний

Количество скрытых состояний	Правдоподобие для скорости двигателя	Правдоподобие для тока двигателя	Правдоподобие для температур двигателя	Правдоподобие для момента двигателя
1	-3209.578	-1395.310	-1205.155	-2194.888
2	-1862.034	-1354.160	-627.526	-1992.406
3	-1471.952	-1350.455	2111.237	-1881.178
4	-929.213	1433.558	3855.816	-1844.285
5	-854.145	2530.130	3680.103	-1831.114
6	1397.477	2805.754	3577.317	-1279.328
7	1730.829	2805.754	3504.390	-156.165
8	2247.008	2805.754	3447.823	-1802.470
9	2247.744	2677.177	3401.605	112.651
10	2326.328	2580.138	3362.527	950.504
11	3431.743	2595.189	3328.677	1019.998
12	3432.063	2504.925	3298.819	1493.333
13	3456.351	2466.610	3272.110	1485.716
14	3053.873	2425.217	3247.949	1309.255

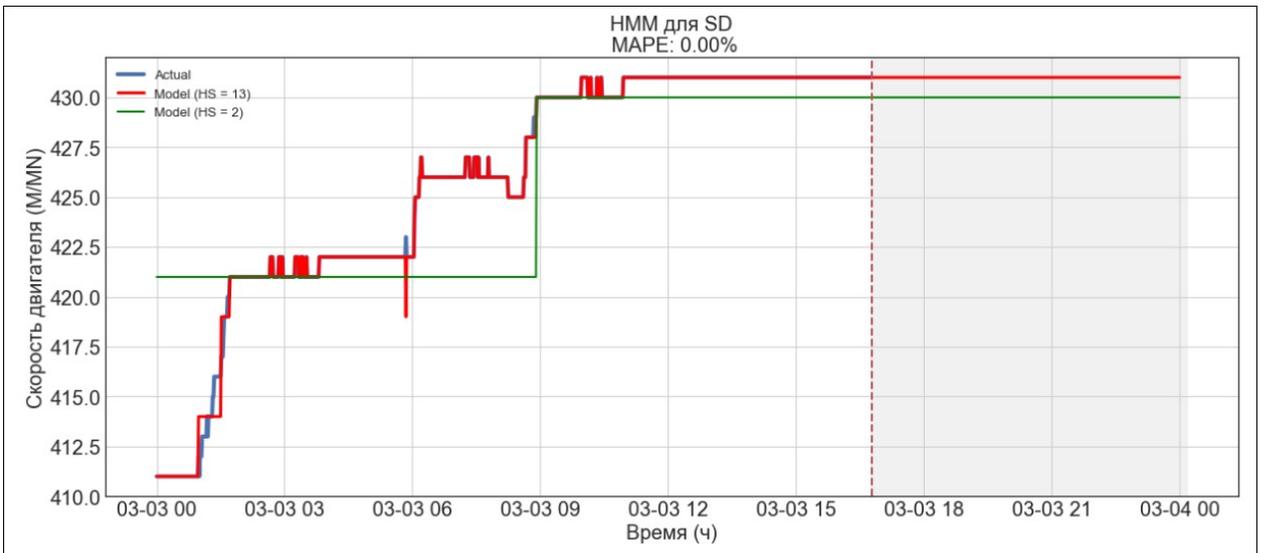


Рисунок 40 – НММ для скорости двигателя

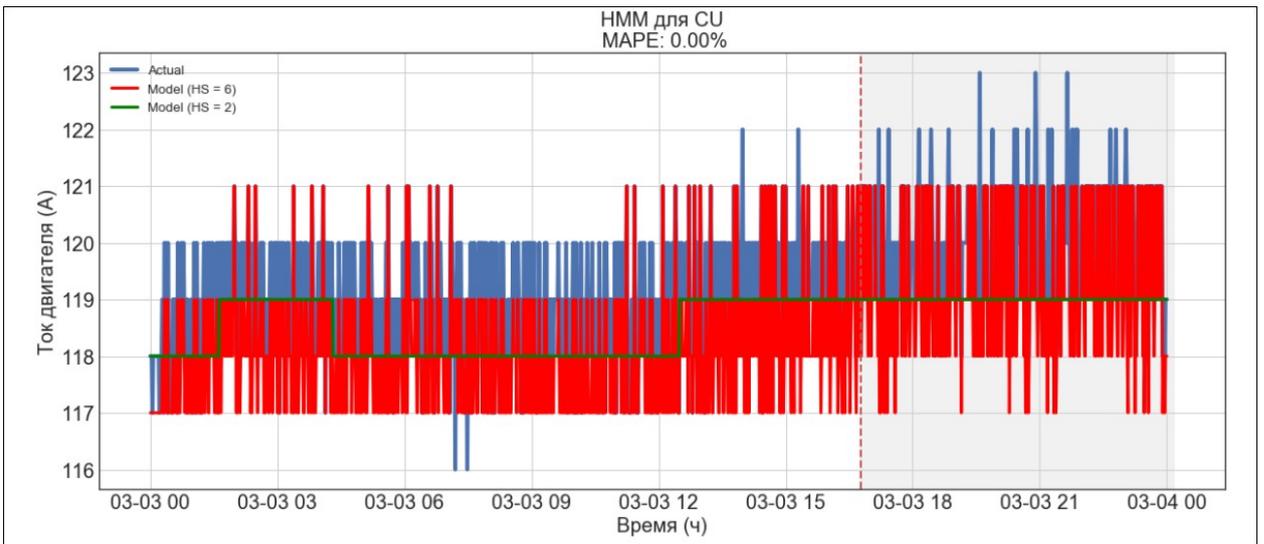


Рисунок 41 – НММ для тока двигателя

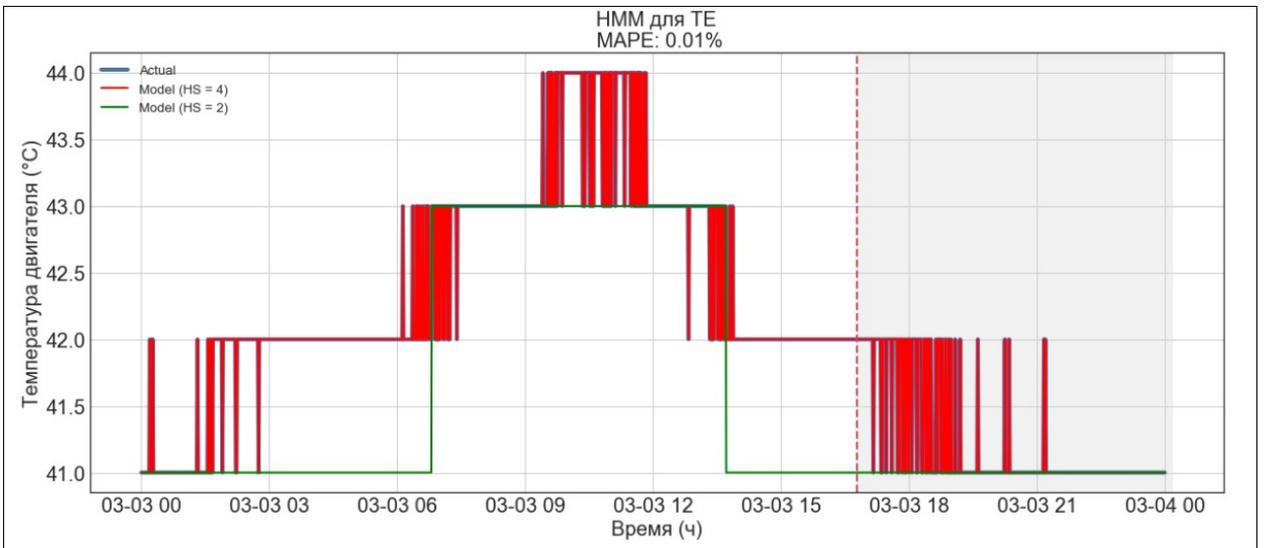


Рисунок 42 – НММ для температуры двигателя

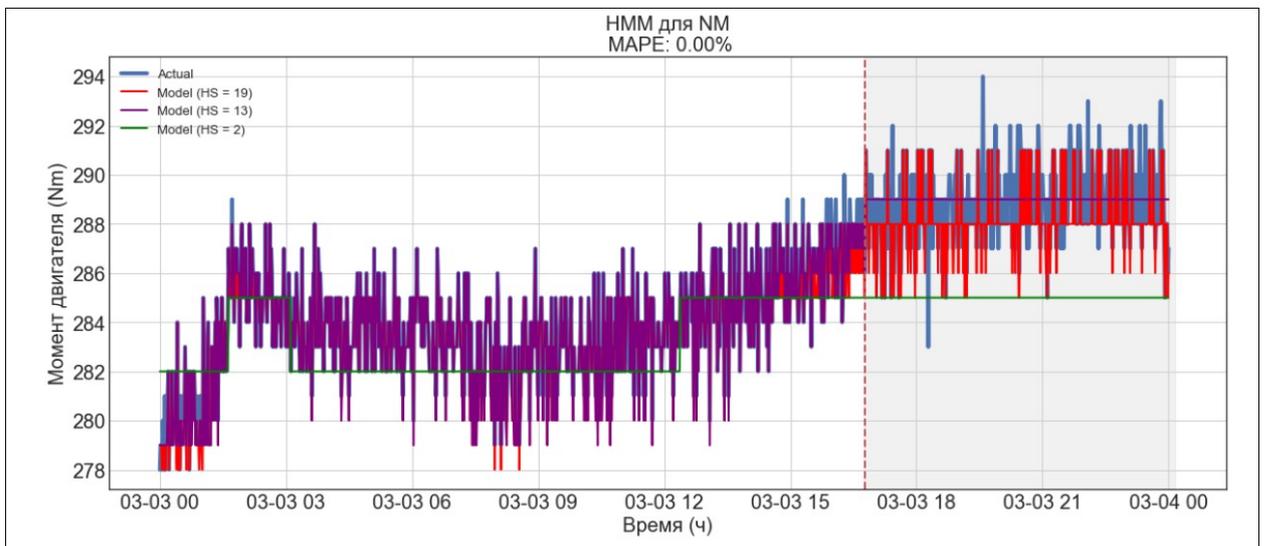


Рисунок 43 – НММ для момента двигателя

Теоретическое обоснование метода представлено в статье [42].

4.8.5 Регрессионные модели

4.8.5.1 Теоретическая часть

Регрессионная модель – функция независимой переменной и параметров с добавленной случайной переменной. Модель сводится к суммированию значений признаков с некоторыми весами:

$$y(x) = w_0 + \sum_{j=1}^d w_j x_j, \quad (18)$$

где w_j – веса модели;

w_0 – свободный коэффициент или сдвиг (bias).

Для обучения регрессионной модели необходимо определить, как измеряется качество предсказаний. Основным способом оценки отклонения прогноза от истинного ответа является *RMSE* (так как *MSE* плохо интерпретируется – не сохраняет единицы измерения):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}. \quad (19)$$

Данная метрика хорошо подходит для сравнения моделей, но не позволяет сделать вывод о том, насколько хорошо модель решает

поставленную задачу, поэтому кроме неё применяется коэффициент детерминации:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n \left(y_i - \sum_{i=1}^n y_i \right)^2}. \quad (20)$$

Коэффициент детерминации измеряет долю дисперсии, объяснённую моделью, в общей дисперсии целевой переменной. Фактически, данная мера качества – нормированная среднеквадратичная ошибка. Если она близка к единице, то модель хорошо объясняет данные.

Исходя из ограничений (см. 4.4) и опираясь на теорему Гаусса-Маркова можно утверждать, что обучение линейной регрессии с использованием среднеквадратичной ошибки – то есть МНК – даёт лучшую оценку параметров модели.

Поиск параметров подразумевает задачу оптимизации, которая решается градиентным бустингом. Градиент является направлением наискорейшего роста функции, а антиградиент – направлением наискорейшего убывания. Соответственно, происходит старт из некоторой точки, далее сдвиг в сторону антиградиента, пересчет антиградиента и снова сдвиг в его сторону – и так повторяется до минимума.

4.8.5.2 Практическая часть

При использовании для прогнозирования авторегрессионных моделей, скрытых марковских моделей и моделей экспоненциального сглаживания использовался одномерный временной ряд. Для линейной регрессии необходимо воспользоваться остальными имеющимися данными и по этим данным предсказывать целевые значения.

После преобразований набор данных имеет следующий вид (рисунок 44.):

	Motor speed	Current	Temperature	Torque
datetime				
2020-03-03 00:00:00	411	118	41	278
2020-03-03 00:01:00	411	118	41	279
2020-03-03 00:02:00	411	118	41	279
2020-03-03 00:03:00	411	117	41	278
2020-03-03 00:04:00	411	118	41	280

Рисунок 44 – Набор признаков для модели линейной регрессии

Данные измерены в разных масштабах, соответственно необходимо произвести нормализацию и привести к общей шкале. Для этого используется стандартная оценка:

$$z = \frac{x - \bar{x}}{S}, \quad (21)$$

где \bar{x} – среднее выборки;

S – стандартное отклонение выборки.

На рисунке 45 представлены прогнозы линейных регрессионных моделей, обученные на трех признаках:

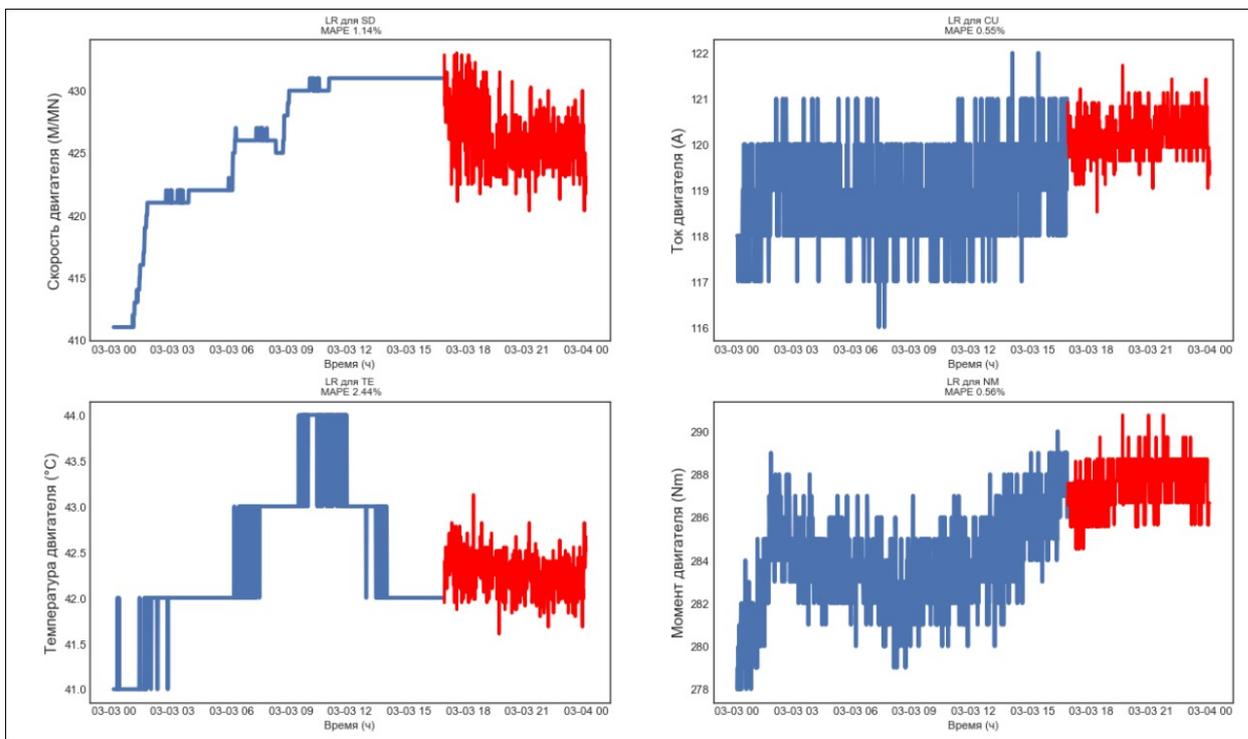


Рисунок 45 – Прогнозирование моделью линейной регрессии

В качестве признаков, кроме характеристик двигателя, можно использовать лаги временного ряда. Сдвинув ряд на n шагов, добавить признак, в котором текущему значению ряда в момент времени t будет

соответствовать его значение в момент времени $t - n$, что позволит делать предсказания на n шагов вперед.

	y	lag_700	lag_701	lag_702	lag_703	lag_704	lag_705	lag_706	lag_707	lag_708	...	lag_990	lag_991	lag_992	lag_993	lag_994	lag_995
datetime																	
2020-03-03 23:54:00	431	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	...	426.0	426.0	426.0	426.0	427.0	427.0
2020-03-03 23:55:00	431	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	...	426.0	426.0	426.0	426.0	426.0	427.0
2020-03-03 23:56:00	431	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	...	426.0	426.0	426.0	426.0	426.0	426.0
2020-03-03 23:57:00	431	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	...	427.0	426.0	426.0	426.0	426.0	426.0
2020-03-03 23:58:00	431	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	...	427.0	427.0	426.0	426.0	426.0	426.0
2020-03-03 23:59:00	431	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	...	427.0	427.0	427.0	426.0	426.0	426.0
2020-03-04 00:00:00	431	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	431.0	...	427.0	427.0	427.0	427.0	426.0	426.0

Рисунок 46 – Набор признаков с лагами временного ряда

Большое количество лагов приводит к переобучению, малое к ухудшению качества прогноза. Предсказания в зависимости от количества лагов представлены на рисунке 47:

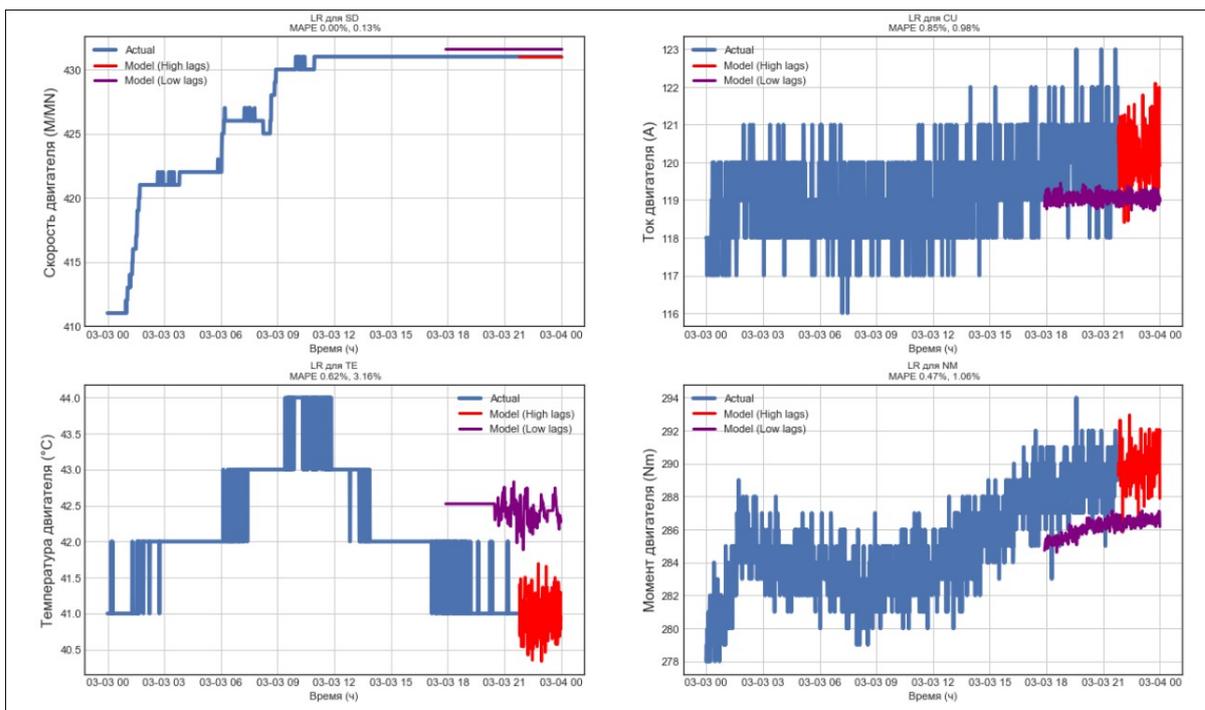


Рисунок 47 – Предсказания с использованием признаков, полученных с помощью лагов временного ряда

Теоретическое обоснование метода представлено в работе [43].

4.8.6 Модели на базе классификационно-регрессионных деревьев

4.8.6.1 CatBoost

В основе библиотеки CatBoost лежит градиентный бустинг на деревьях решений. В ней поддержано несколько видов деревьев, в том числе «симметричные» деревья, которые используются в библиотеке по умолчанию. Алгоритм работает следующим образом – для каждого параметра имеется набор признаков, имеется дерево, в вершинах дерева – условия. Если условие выполнено осуществляется переход в левый узел, в противном случае в правый. При построении дерева необходимо задать функционал качества, на основе которого осуществляется разбиение выборки на каждом шаге.

$$Q(R_m, j, s) = H(R_m) - \frac{|R_l|}{|R_m|} H(R_l) - \frac{|R_r|}{|R_m|} H(R_r), \quad (22)$$

R_l и R_r – объекты, попадающие в левое и правое поддерево;

$H(R)$ – критерий информативности, который оценивает качество распределения целевой переменной.

Критерий информативности в случае регрессии переписывается в следующем виде:

$$QH(R_m) = \frac{1}{|R|} \sum_{(x_i, y_i) \in R} (y_i - \frac{1}{|R|} \sum_{(x_i, y_i) \in R} y_i)^2. \quad (23)$$

Информативность вершины измеряется её дисперсией – чем ниже разброс целевой переменной, тем лучше вершина.

Цикл обучения CatBoost состоит из построения одного дерева, проверка качества и переобучения, после чего сохранения результата (рисунок 48.):

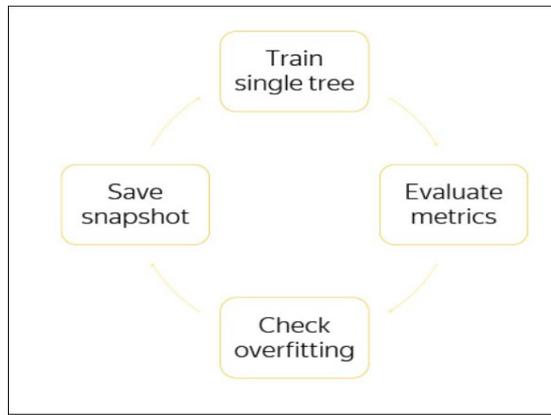


Рисунок 48 – Главный цикл обучения CatBoost [35]

Подготовка данных для предсказаний с помощью CatBoost осуществляется таким же образом, что и для модели линейной регрессии (см.4.8.5.2).

На рисунке 49 представлены прогнозы моделью CatBoost, обученной на данных двигателя, на рисунке 50, на лагах временного ряда:

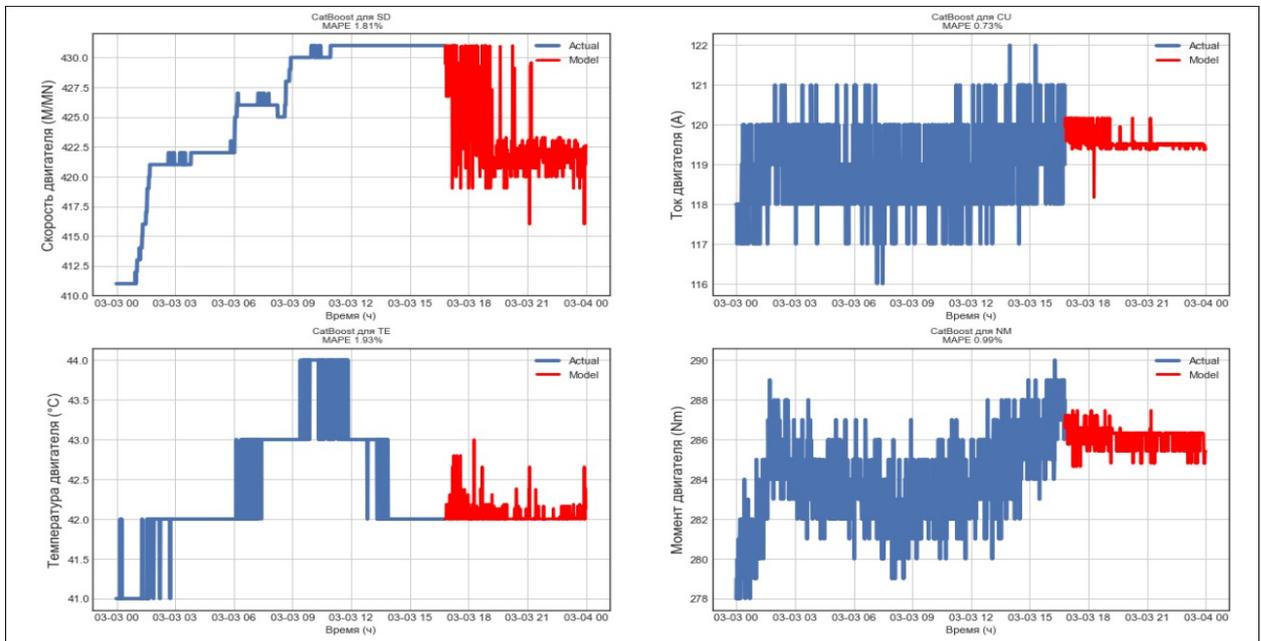


Рисунок 49 – Прогнозирование моделью CatBoost

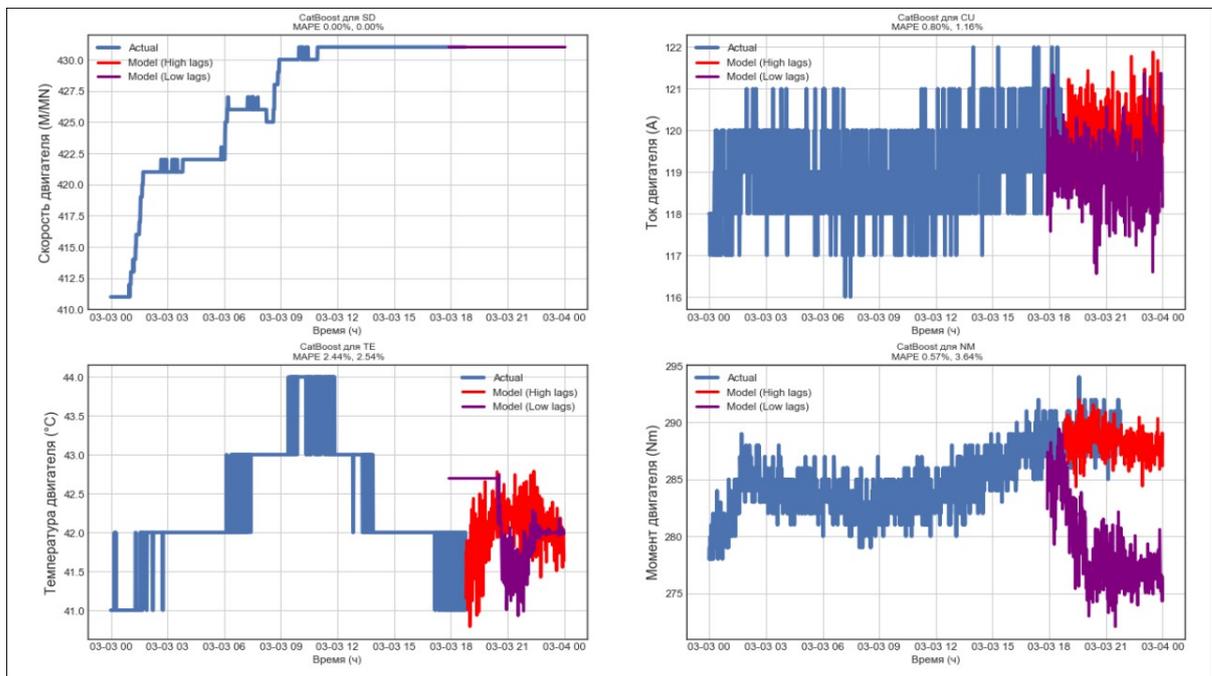


Рисунок 50 – Прогнозирование с использованием признаков, полученных с помощью лагов временного ряда

Теоретическое обоснование метода представлено в работе [44].

4.8.6.2 XGBoost

В основе XGBoost, как и в основе CatBoost, лежит алгоритм градиентного бустинга деревьев решений. Модель предсказания представлена в форме ансамбля слабых предсказывающих моделей – деревьев решений. Обучение ансамбля проводится последовательно. На каждой итерации вычисляются отклонения предсказаний уже обученного ансамбля на обучающей выборке. Следующая модель, которая будет добавлена в ансамбль будет предсказывать эти отклонения. Таким образом, добавив предсказания нового дерева к предсказаниям обученного ансамбля, уменьшается среднее отклонение модели, которое является целью оптимизационной задачи. Новые деревья добавляются в ансамбль до тех пор, пока ошибка уменьшается, либо пока не выполняется одно из правил "ранней остановки".

Особенности реализации алгоритма в XGBoost заключается в использовании помимо первой производной – второй производной от

функции потерь, наличие встроенной регуляризации, возможность задавать пользовательские функции потерь и метрики качества.

Подготовка данных для предсказаний с помощью XGBoost осуществляется таким же образом, что и для модели линейной регрессии (см.4.8.5.2) и для CatBoost.

На рисунке 51 представлены прогнозы моделью XGBoost, обученной на данных двигателя, на рисунке 52, на лагах временного ряда:

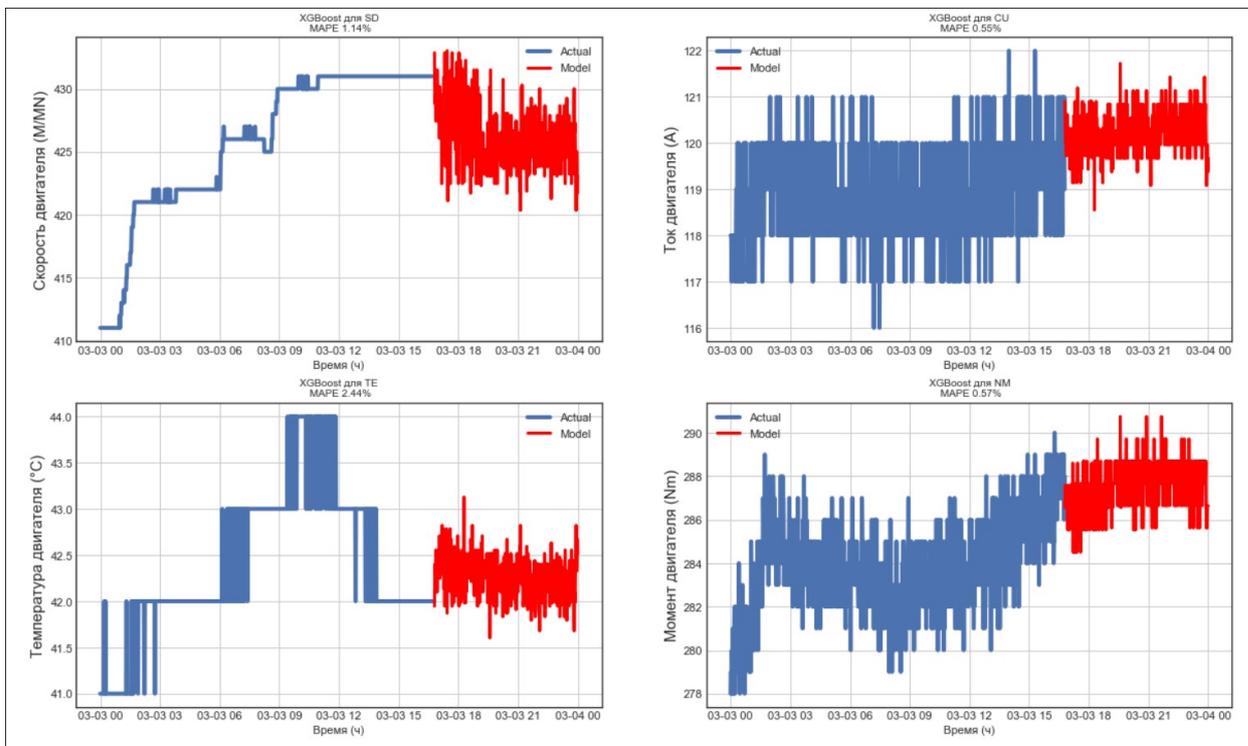


Рисунок 51 – Прогнозирование моделью XGBoost

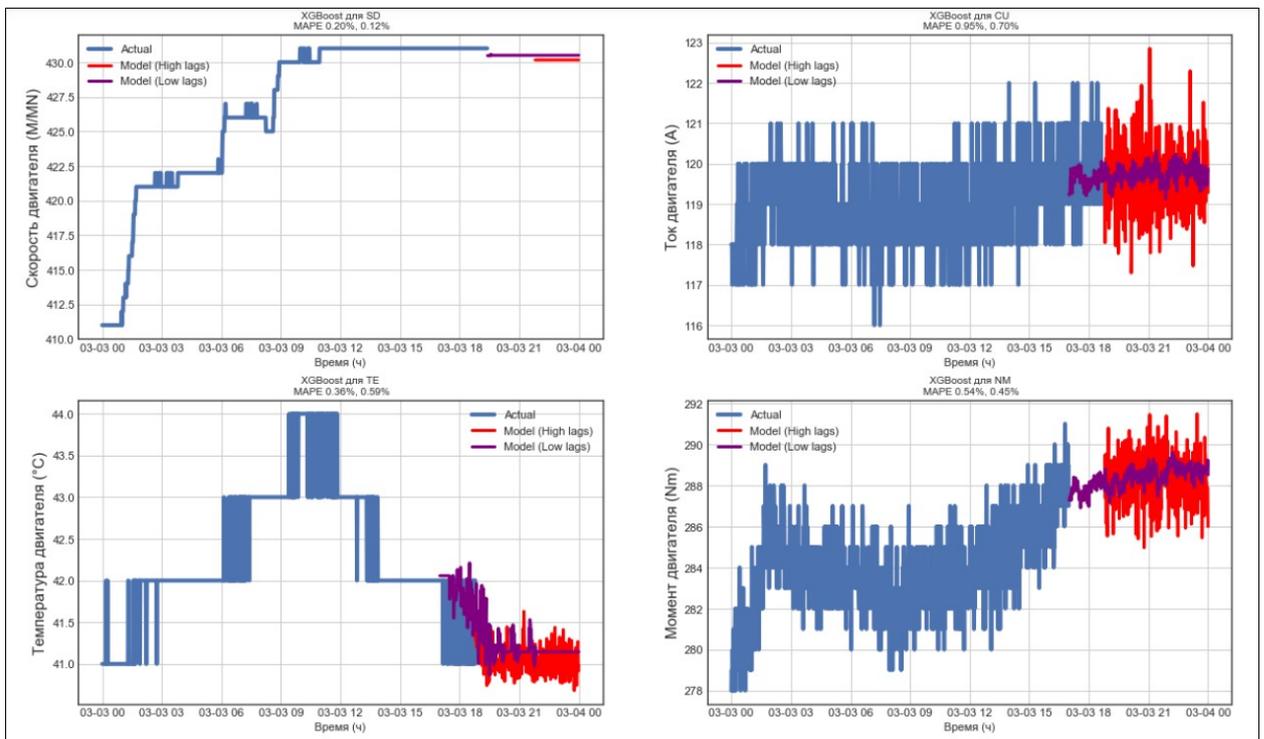


Рисунок 52 – Прогнозирование с использованием признаков, полученных с помощью лагов временного ряда

Теоретическое обоснование метода представлено в работе [45].

4.8.6.3 LightGBM

LightGBM — фреймворк от Microsoft, основное преимущество которого состоит в скорости обучения на больших массивах данных. В основе него, как и в случае CatBoost'a, и XGBoost'a лежит алгоритм градиентного бустинга деревьев решений. Отличие в обучении приведена на рисунке 54.

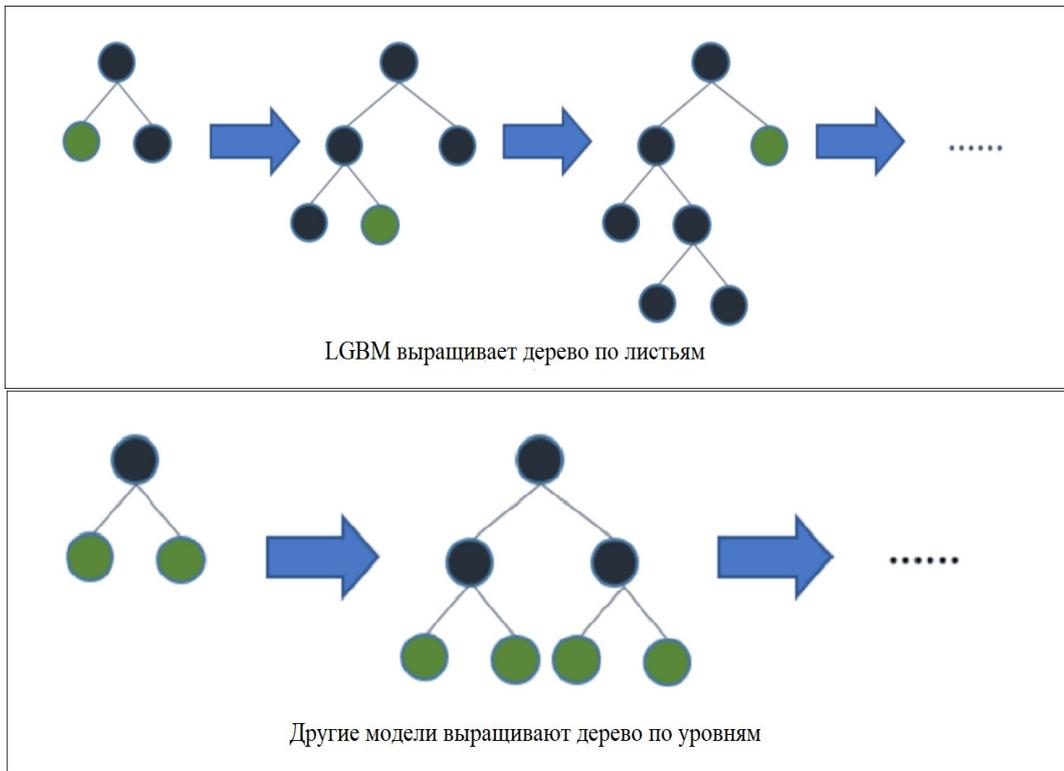


Рисунок 53 – Отличия алгоритма обучения LGBM от других моделей

Подготовка данных для предсказаний с помощью LGBM осуществляется таким же образом, что и для модели линейной регрессии (см.4.8.5.2), для CatBoost и для XGBoost.

На рисунке 54 представлены прогнозы моделью LGBM, обученной на данных двигателя, на рисунке 55, на лагах временного ряда.

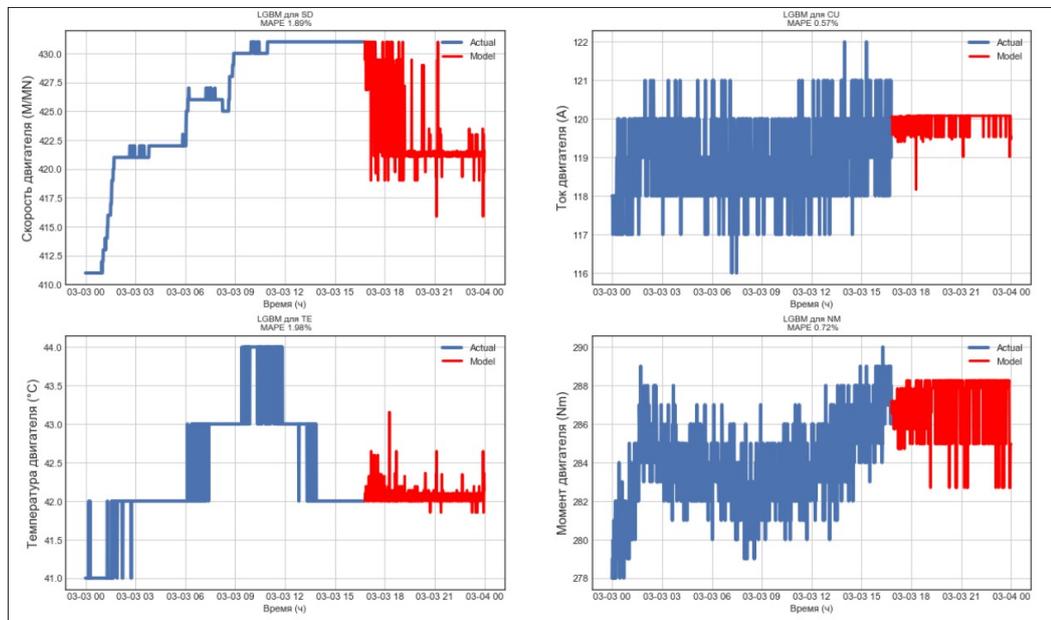


Рисунок 54 – Прогнозирование моделью LGBM

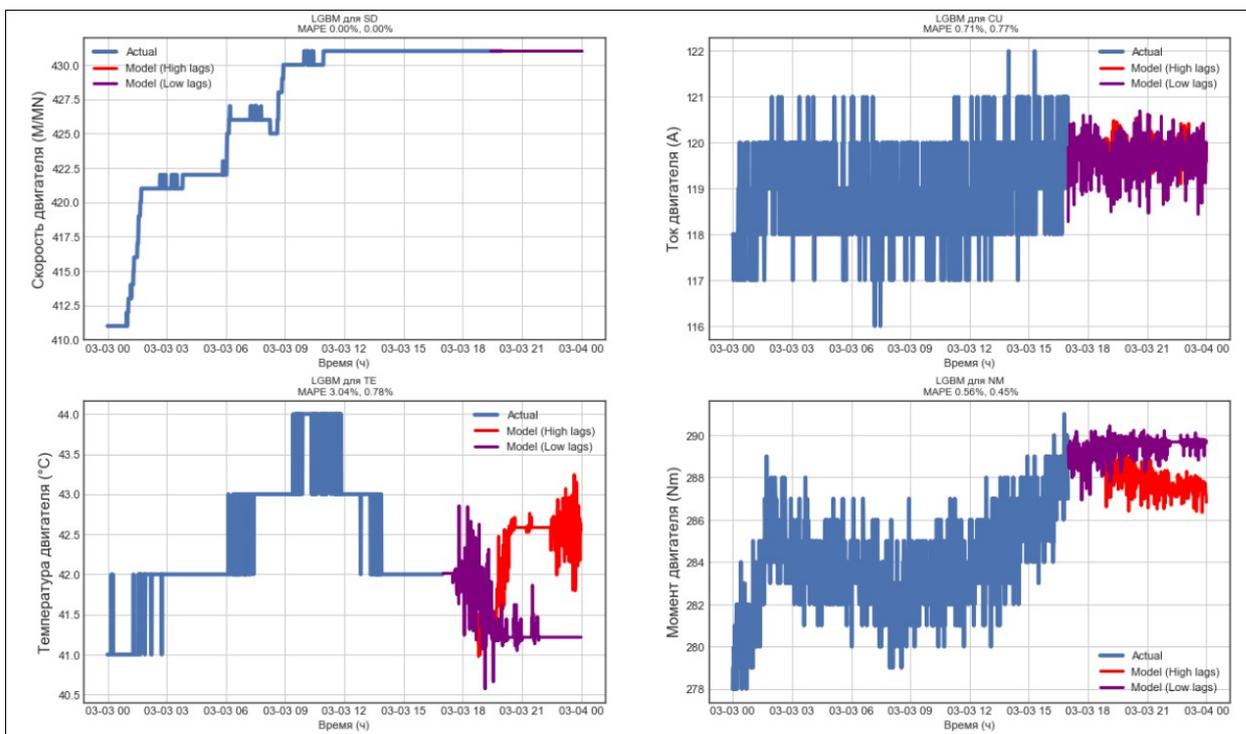


Рисунок 55 – Прогнозирование с использованием признаков, полученных с помощью лагов временного ряда

Теоретическое обоснование метода представлено в работе [46].

4.8.6.4 Сравнение CatBoost, XGBoost, LightGBM

CatBoost, в отличие от других алгоритмов, позволяет работать с категориальными данными, не переводя их в числовые значения, что позволяет проще обрабатывать текстовые логи событий производственной линии. Не требует длительной настройки и показывает высокое качество обучения сразу же при применении.

LightGBM использует новую технику односторонней выборки на основе градиента (GOSS) для фильтрации экземпляров данных для нахождения значения разделения, в то время как XGBoost использует предварительно отсортированный алгоритм и алгоритм на основе гистограммы для вычисления наилучшего разделения.

В таблице 7 приведены различия в гиперпараметрах алгоритмов.

Таблица 7 – Гиперпараметры алгоритмов

Задача гиперпараметра	XGBoost	CatBoost	Light GBM
Параметры, контролирующие переобучение	1. learning_rate 2. max_depth 3. min_child_weight	1.learning_rate 2.depth 3.l2-leaf-reg	1.learning_rate 2.max_depth 3.num_leaves 4.mid_data_in_leaf
Параметры для категориальных переменных	–	1.cat_features 2.one_hot_max_size	1.categorical_feature
Параметры для контроля скорости обучения	1.colsample_bytree 2.subsample 3.n_estimators	1.rsm 2.iterations	1.feature_fraction 2.bagging_fraction 3.num_iterations

Так как CatBoost может обрабатывать категориальные данные, его настройка не нуждается в большом количестве времени, имеет высокую точность и скорость, то применительно к производству и потоковым временным рядам из классификационно-регрессионных алгоритмов следует выбрать его на стадии интеграции в технологический процесс.

4.8.7 Нейросетевые модели

Для прогнозирования временных рядов используются рекуррентные нейронные сети (Recurrent Neural Network, RNN). Рекуррентная нейронная сеть не является сетью прямого распространения: она содержит циклы. Каждый узел u рекуррентного слоя l имеет вещественное состояние $h_{l,u}$. Состояние рассматривается как память узла. В RNN каждый узел u в каждом слое l имеет два входа: вектор состояний из предыдущего слоя $l-1$ и вектор состояний из этого же слоя l , но из предыдущего временного шага.

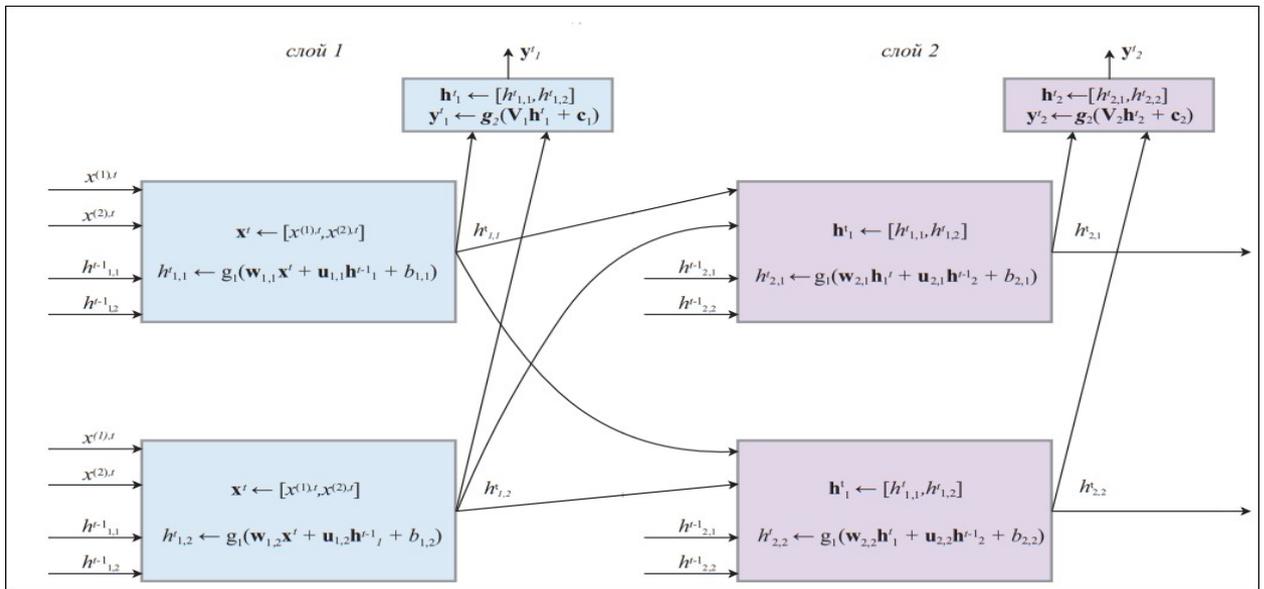


Рисунок 56 – Первые два слоя в рекуррентной нейронной сети

На рисунке 56 каждый обучающий образец представлен матрицей, в которой каждая строка является вектором признаков. Векторы признаков из входных параметров последовательно обрабатываются нейронной сетью в порядке временных шагов. Для обновления состояния $h_{l,u}^t$ вычисляется линейная комбинация входного вектора признаков с вектором состояния того же слоя из предыдущего временного шага. Значение $h_{l,u}^t$ получается применением функции активации \tanh к результату линейной комбинации. Для вычисления выхода y_i^t применяется функция активации softmax . Значения $w_{l,u}$, $u_{l,u}$, $b_{l,u}$, $V_{l,u}$, $c_{l,u}$ определяются по обучающим данным с использованием градиентного спуска с обратным распространением.

Функции \tanh и softmax имеют проблему затухания градиента, кроме того, RNN страдает проблемой в обработке долгосрочных зависимостей. По мере увеличения длины входной последовательности векторы признаков, находящиеся в начале последовательности, постепенно «забываются». Поэтому применяются более эффективная архитектура рекуррентной нейронной сети – сети с долгой краткосрочной памятью (Long Short-Term Memory, LSTM).

LSTM разработаны специально, чтобы избежать проблемы долговременной зависимости. Структура LSTM представлена на рисунке 58:

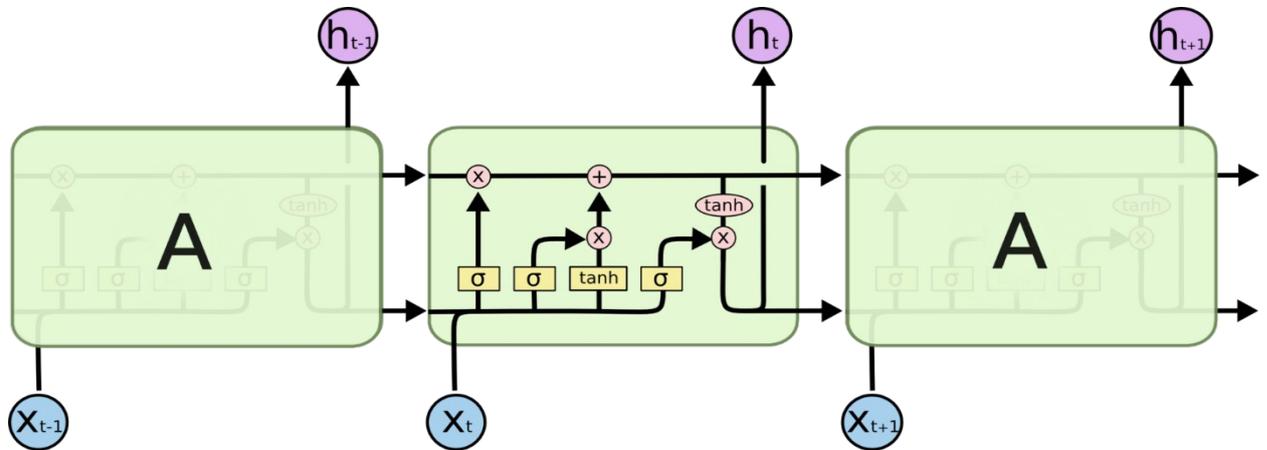


Рисунок 57 – Структура LSTM

В качестве входных данных модели используется одномерный временной ряд. Количество эпох для обучения выбрано 2 и 50, размер батча 1. В качестве метода оптимизации используется «Adam» [47]. Предсказания представлены на рисунках 58, 59, 60, 61.

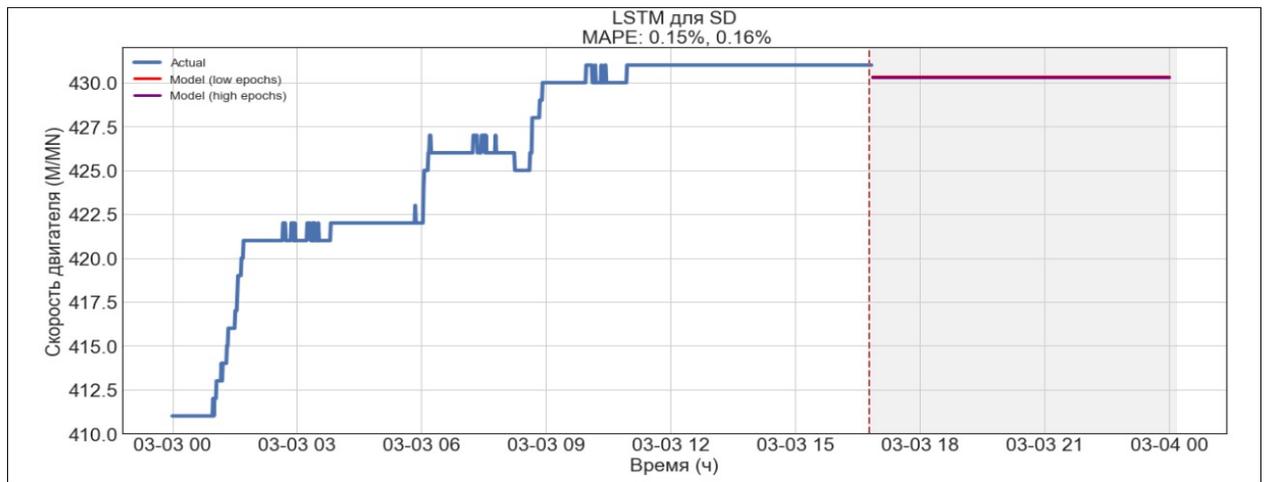


Рисунок 58 – Прогнозирование моделью LSTM для SD

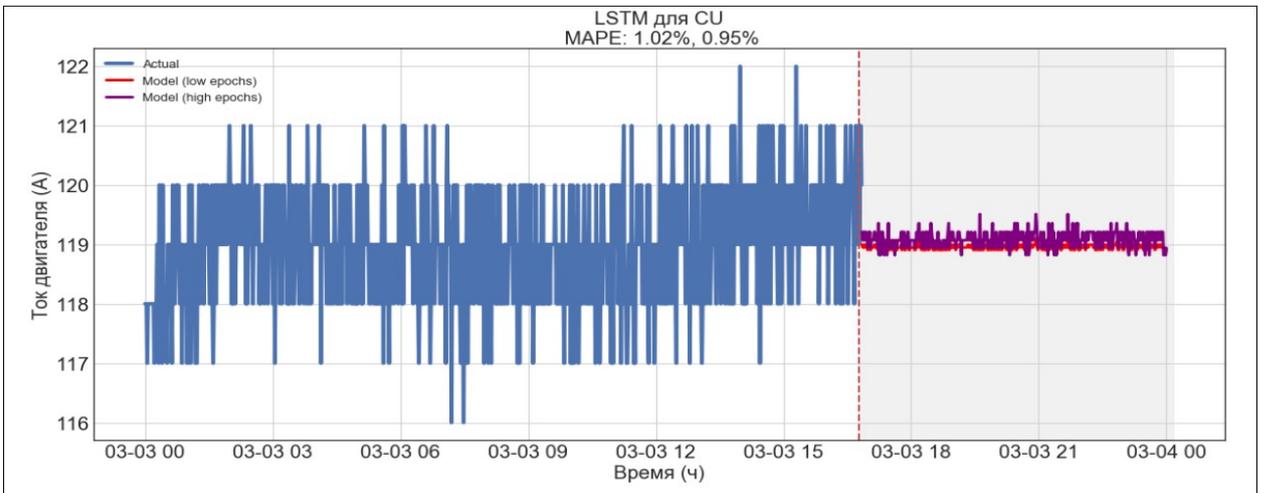


Рисунок 59 – Прогнозирование моделью LSTM для CU

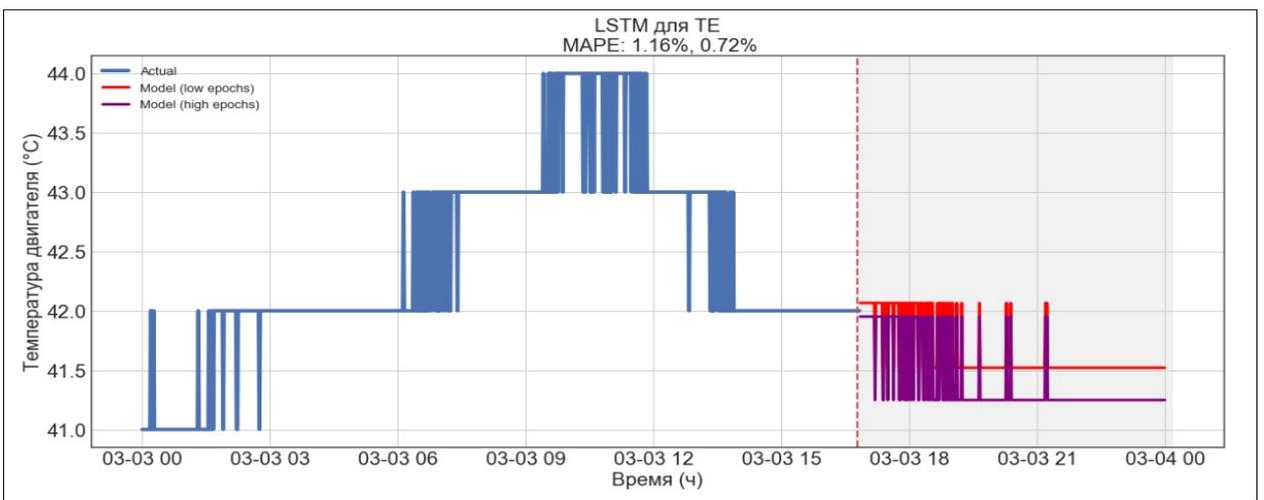


Рисунок 60 – Прогнозирование моделью LSTM для TE

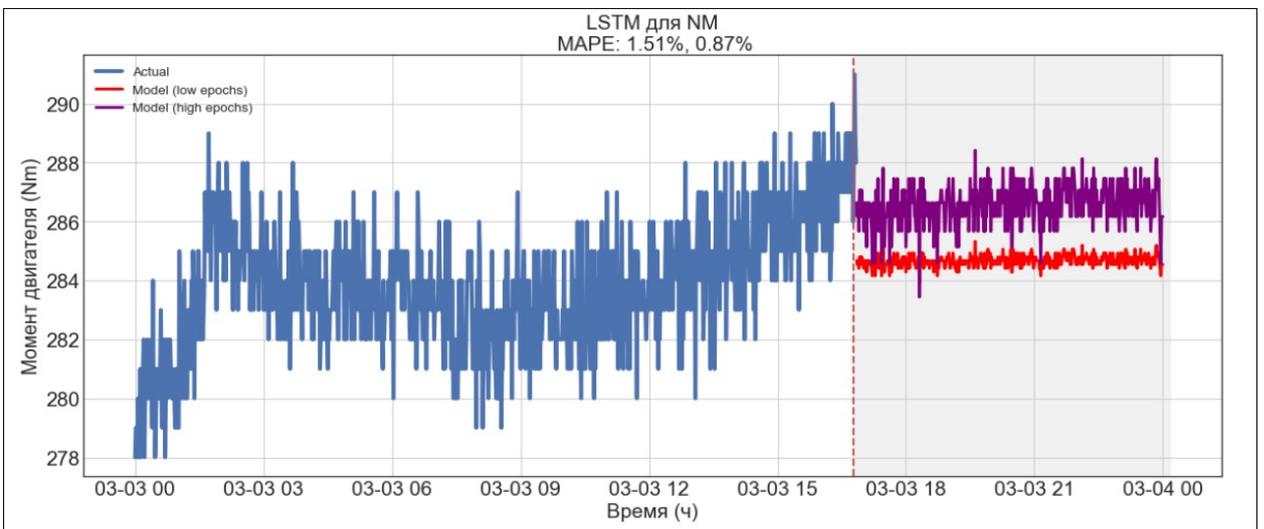


Рисунок 61 – Прогнозирование моделью LSTM для NM

Теоретическое обоснование модели представлено в работе [48].

4.8.8 Готовые решения (Facebook Prophet, Amazon Gluon TS)

4.8.8.1 Facebook Prophet

Facebook Prophet – библиотека разработанная Facebook для прогнозирования временных рядов. Позволяет быстро и достаточно качественно автоматически строить прогнозы. Модель, на основе которой работает Facebook Prophet является additive regression model, состоящая из следующих компонент:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t, \quad (24)$$

где $s(t)$ – сезонная компонента отвечающая за моделирование периодических изменений, связанных с недельной и годовой сезонностью; $g(t)$ – тренд, кусочно-линейная или логистическая функция;

$h(t)$ – отвечает за аномальные дни;

ε_t – ошибка содержит информацию, которая не учтена моделью.

В качестве входных данных используется одномерный временной ряд.

Прогнозы представлены на рисунке 62.

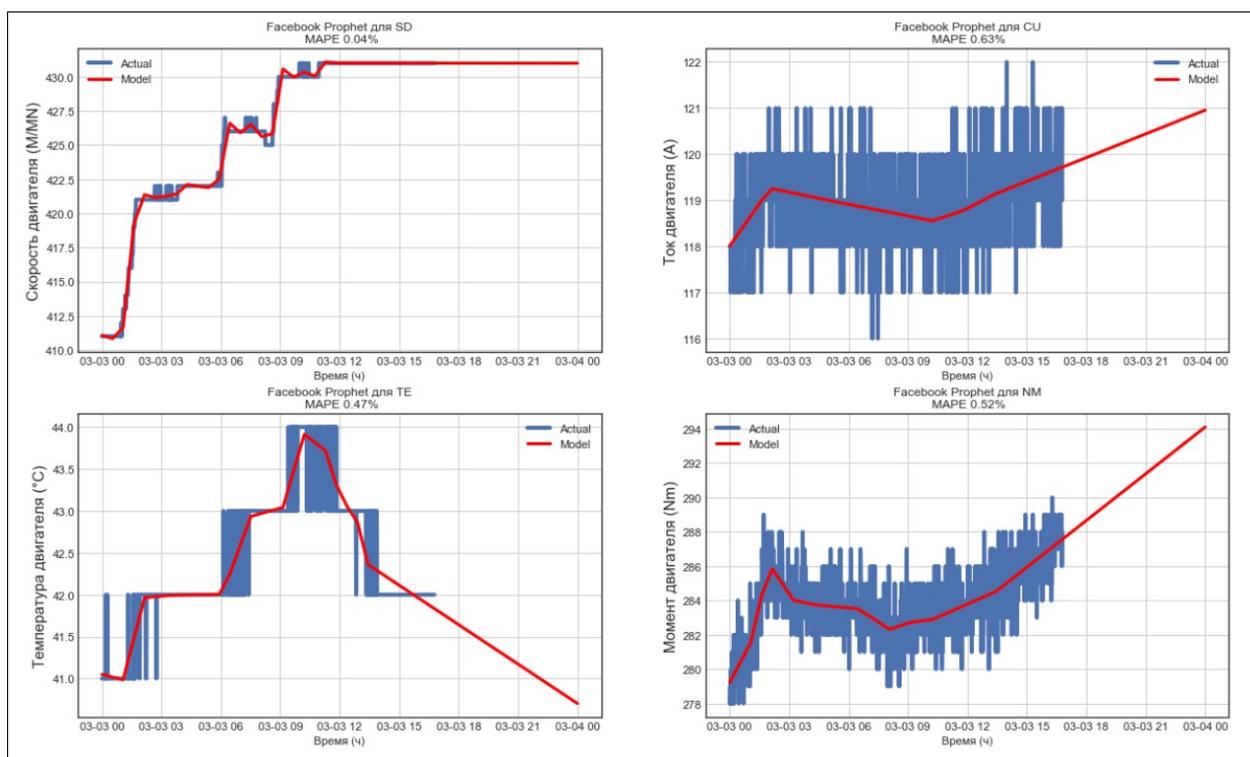


Рисунок 62 – Прогнозирование Facebook Prophet

Теоретическое обоснование метода представлено в работе [49].

4.8.8.2 Amazon Gluon TS

Amazon Gluon TS – библиотека, разработанная компанией Amazon, позволяющая быстро и эффективно строить, оценивать и использовать модели временных рядов, основанные на глубоком обучении и вероятностном подходе. Amazon использовал рекуррентные нейронные сети (LSTM), а также свертки и механизмы внимания. Gluon TS содержит:

1. инструменты, необходимые для построения и обучения наиболее распространённых архитектур нейронных сетей, а также компоненты для моделирования и трансформации вероятностных распределений;
2. механизмы для загрузки и предварительной обработки данных, в том числе автоматической генерации признаков из временных рядов;
3. несколько готовых к использованию state-of-the-art прогнозных моделей;
4. инструменты для оценки и сравнения различных моделей.

В качестве входных использовался одномерный временной ряд.

В качестве модели используется архитектура, основанная на глубокой авторегрессионной модели (DeerAR). Прогнозы представлены на рисунке 63, 64, 65, 66.

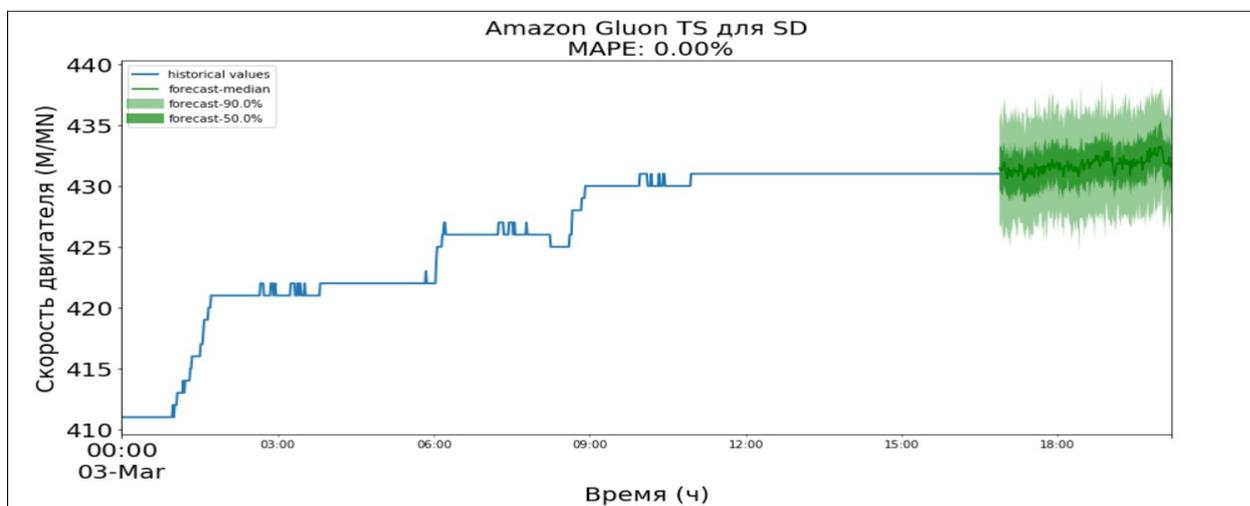


Рисунок 63 – Прогнозирование Amazon Gluon TS для SD

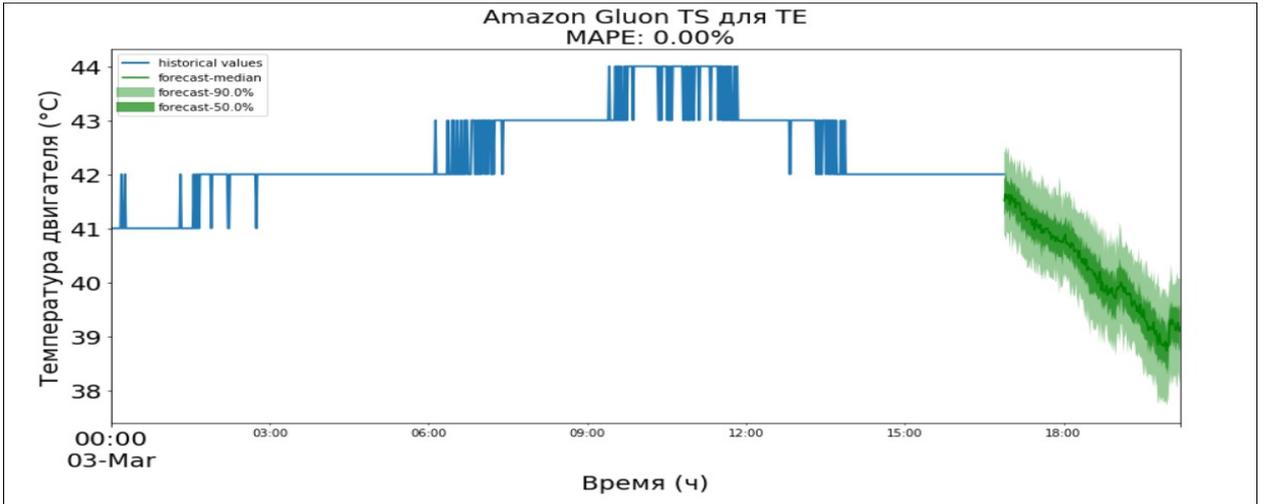


Рисунок 64 – Прогнозирование Amazon Gluon TS для TE

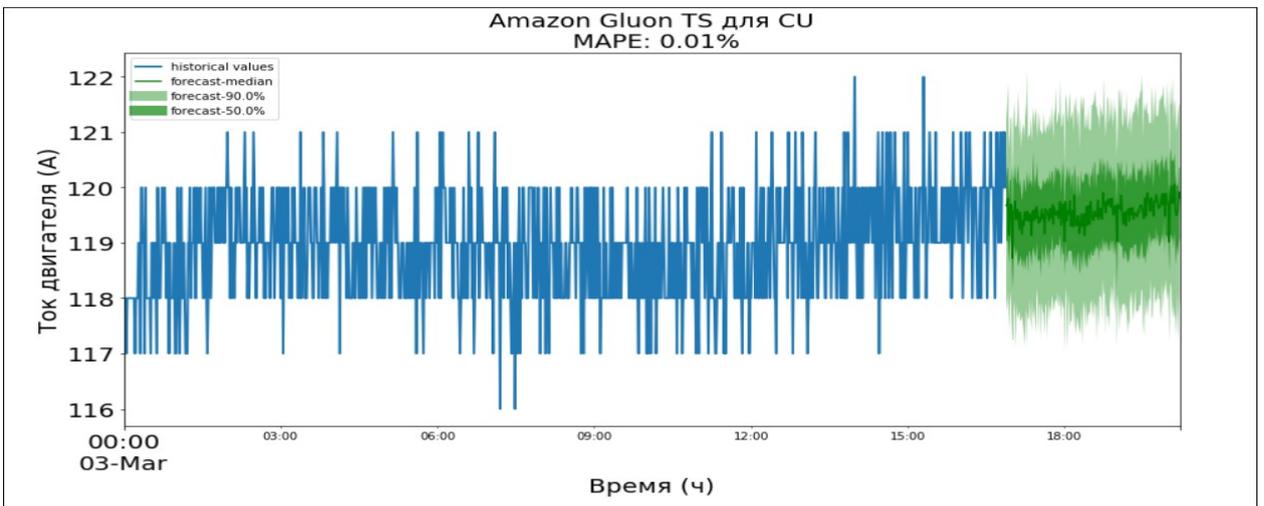


Рисунок 65 – Прогнозирование Amazon Gluon TS для CU

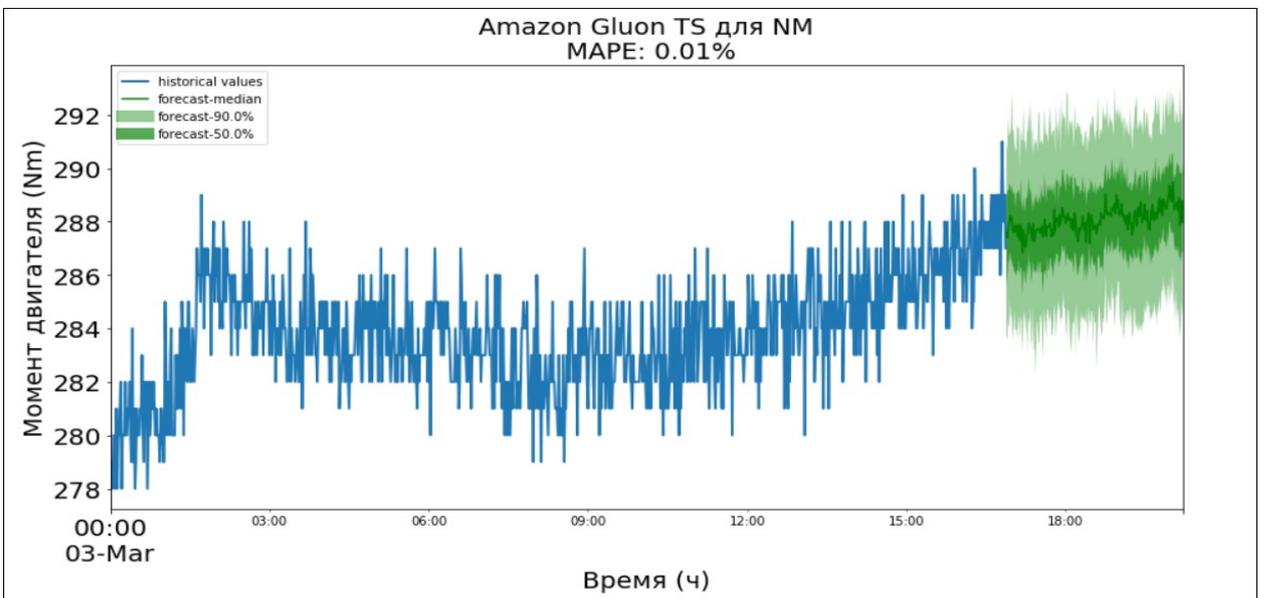


Рисунок 66 – Прогнозирование Amazon Gluon TS для NM

Теоретическое обоснование метода представлено в работе [50].

4.8.9 Сравнение моделей прогнозирования

В предыдущих разделах рассмотрены основные модели прогнозирования, их математический аппарат и способы применения. Получены прогнозы на несколько часов вперед каждого параметра двигателя, а также рассчитана средняя абсолютная процентная ошибка для каждого из прогнозов.

В таблице 8 приведены достоинства и недостатки каждой из моделей.

Таблица 8 – Сравнение моделей

Модель	Достоинства	Недостатки
Модели экспоненциального сглаживания	Наглядность и интерпретируемость результатов; учет весов исходной информации; простота вычислительных операций.	В случае простого и двойного ЭС короткие шаги прогнозирования, в случае тройного ЭС подбор параметров; недостаточная гибкость в описании динамики процессов.
Авторегрессионные модели	Наглядность и интерпретируемость результатов; высокая гибкость к динамике процесса; простота вычислительных операций.	Низкая адаптивность; неоднозначное влияние параметров на построение модели; затратный подбор оптимальных параметров; для AR, MA, ARMA требуется стационарный временной ряд.
Линейные регрессионные модели	Наглядность и интерпретируемость результатов; простота вычислительных операций.	Недостаточная точность прогноза на более сложных динамических процессах; большое количество исходных данных.

Продолжение таблицы 8 – Сравнение моделей

Модель	Достоинства	Недостатки
Модели на базе цепей Маркова	Высокая скорость моделирования; простая математическая структура; моделирование сложной цепочки наблюдений.	Проблема выбора алгоритма прогнозирования, выбора закона распределения; не учитывает зависимости между соседними элементами временного ряда.
Модели на базе классификационно-регрессионных деревьев	Гибкость к динамике процесса; масштабируемость; возможность учитывать категориальные переменные.	Подбор гиперпараметров модели; количество входных данных; неоднозначность алгоритмам.
Нейросетевые модели	Гибкость к динамике процесса; масштабируемость.	Низкая скорость; подбор гиперпараметров; неоднозначность алгоритма.
Facebook Prophet, Amazon Gluon TS	Гибкость к динамике процесса; масштабируемость; простота использования.	Низкая скорость; подбор гиперпараметров; низкая интерпретируемость.

Программный код одного из методов прогнозирования приведен в приложении Б.

4.9 Детектирование аномалий

Одной из подзадач предиктивной модели является обнаружение аномальных значений. Аномалия – некоторое отклонение производственного процесса от обычного для него течения. Исходя из обнаруженных аномалий можно судить о том, что процесс нестабилен, оборудование выходит из строя и т.д.. Аномалия может быть сигналом к действию, к проведению проверки, к неправильной эксплуатации, смене режимов работы. Обнаруженные аномалии могут дать дополнительную информацию в локализации проблемы нарушения хода технологического процесса. Кроме того, поиск аномалий, как экстремальных значений, важен для обучения прогнозной модели своим исключением из набора данных.

Виды аномалий:

1. Выброс – элемент маломощного подмножества выборки, существенно отличающийся от остальных элементов выборки.
2. Выброс временного изменения – постепенно угасающий выброс и ряд возвращается к обычному уровню.
3. Выброс сдвига уровня – выброс, оказывающий перманентное воздействие на все последующие элементы выборки.

Методы детектирования аномалий представлены на рисунке 67 [51].



Рисунок 67 – Методы детектирования аномалий

Аномалии анализируются на примере следующего отрезка временного ряда (2400 наблюдений), представленного на рисунке 68:



Рисунок 68 – Интервал временного ряда скорости двигателя

4.9.1 Статистические методы

4.9.1.1 Ящик с усами

Ящик с усами или диаграмма размаха (рисунок 69.) – графическое представление распределения непрерывных данных. Диаграмма показывает пять характеристик: медиану, первую и третью квантили (межквартильный размах), минимум и максимум. Исходя из неё можно определить выбросы (экстремальные значения) в данных. Но экстремальные значения не всегда являются аномалиями.

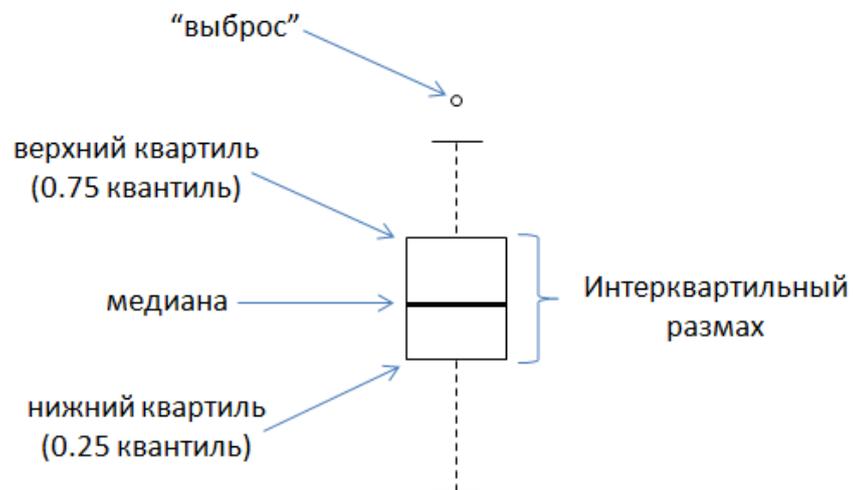


Рисунок 69 – Ящик с усами

Ящик с усами анализируемого временного ряда представлен на рисунке 70:

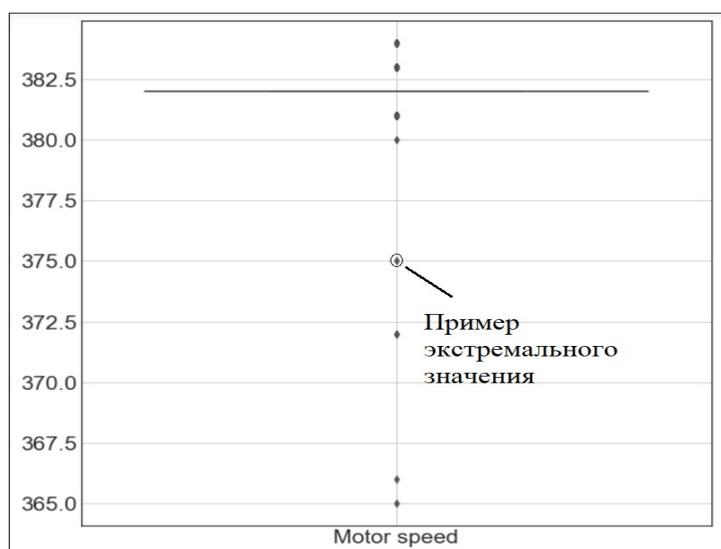


Рисунок 70 – Ящик с усами SD

Теоретическое обоснование метода представлено в работе [52].

4.9.1.2 Z-score

Стандартизованная оценка или z-score – мера относительного разброса наблюдаемого или измеренного значения, которая показывает, сколько стандартных отклонений составляет его разброс относительно среднего значения:

$$z = \frac{x - \bar{X}}{S_x}, \quad (25)$$

где \bar{X} – среднее значение;

S_x – стандартное отклонение для множества данных x_i .

Из правила трех сигм следует, что вероятность отклонения случайного наблюдения от математического ожидания более чем на три среднеквадратических отклонения, практически равна нулю. Соответственно наблюдения, для которых z больше трех будут считаться аномалиями (выбросами/экстремальными значениями).

Аномалии, найденные с помощью стандартизированной оценки, представлены на рисунке 71:

datetime	Motor speed	Motor speed z-score	Anomalies(outliers)
2020-03-10 04:53:00	372	-13.567700	1
2020-03-10 04:54:00	365	-22.876965	1
2020-03-10 04:55:00	366	-21.547070	1
2020-03-10 04:56:00	372	-13.567700	1
2020-03-10 04:57:00	375	-9.578015	1

Рисунок 71 – Аномалии методов z-score для SD

Теоретическое обоснование метода представлено в работе [53].

4.9.2 Контрольные карты

Контрольная карта Шухарта – визуальный инструмент, график изменения параметров технологического процесса во времени. Используется для обеспечения статистического контроля стабильности процесса, что включает в себя и определение аномалий. Карта Шухарта имеет центральную линию (CL), соответствующую эталонному значению характеристики. Две статистические определяемые контрольные границы относительно центральной линии – верхняя контрольная граница (UCL) и нижняя контрольная граница (LCL). Верхние и нижние контрольные границы строятся на расстоянии трех сигм от центральной линии. Помимо этого, строятся границы на расстоянии двух сигм, называемые «предупреждающими», которые служат для предупреждения о приближающемся выходе процесса из состояния статистической управляемости.

Контрольные карты делятся в зависимости от типа данных. В случае рассматриваемых данных выбирается карта выборочных средних, исходя из рисунка 2 ГОСТа [52].

Карта выборочных средних анализируемого временного ряда представлена на рисунке 72:

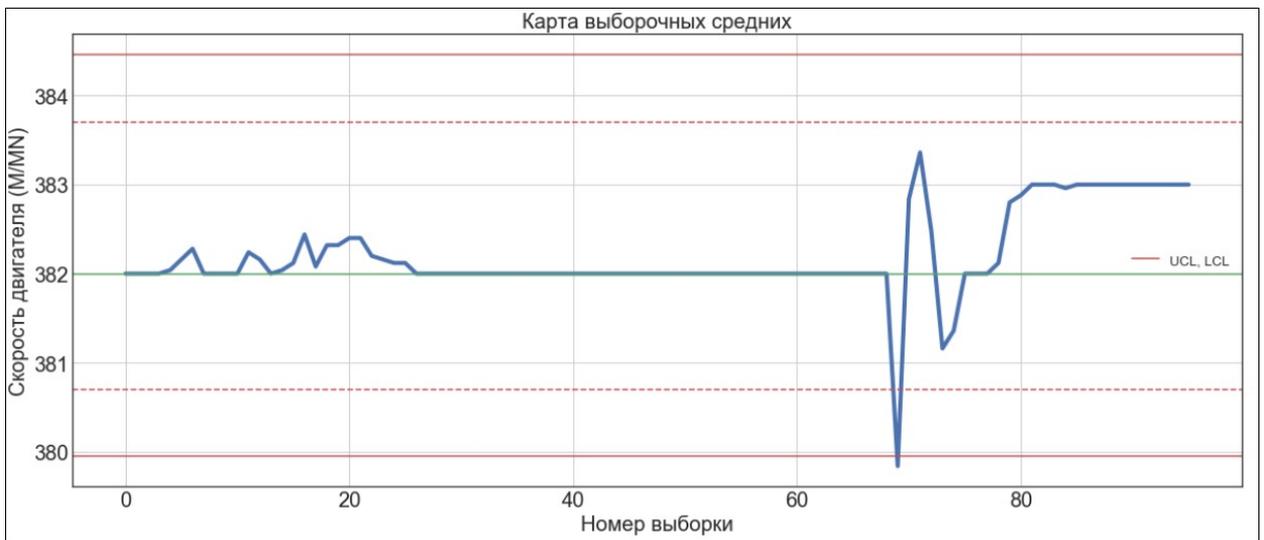


Рисунок 72 – Карта выборочных средних

Теоретическое обоснование метода представлено в работе [54].

4.9.3 Методы, основанные на плотностях

4.9.3.1 Локальный уровень выброса (LOF, Local outlier factor)

Локальный уровень выброса классифицирует аномалии с использованием локальной плотности около некоторого элемента выборки. Локальная плотность точки данных – это концентрация других точек в ближайшей окрестности этой точки, где размер такой окрестности может быть определен либо как постоянное пороговое значение расстояния, либо по ближайшим n соседним точкам. LOF численно оценивает изолированность одной точки данных по отношению к ее ближайшим n соседям. Точки данных с существенно более низкой локальной плотностью по сравнению с n соседями считаются аномалиями. Расстояние до аномалии рассчитывается специальными метриками, представленными в таблице 9:

Таблица 9 – Метрики расстояний

Метрика расстояния	Формула
Евклидово расстояние	$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$
Расстояние Чебышева	$\max_i u_i - v_i $

Продолжение таблицы 9 – Метрики расстояний

Метрика расстояния	Формула
--------------------	---------

Расстояние Махаланобиса	$\sqrt{(u-v)V^{-1}(u-v)^T}$
Расстояние Минковского	$(\sum_{i=1}^n (w_i(u_i-v_i) ^p))^{1/p}$

Применительно к исследуемому объекту использовалось расстояние Минковского, количество соседей 20. Найденные аномалии классифицируются меткой «-1», представлены на рисунке 73:

Date	Motor speed	Anomalies	Date	Motor speed	Anomalies
2020-03-10 04:53:00	372	-1	2020-03-10 05:31:00	384	-1
2020-03-10 04:54:00	365	-1	2020-03-10 05:32:00	384	-1
2020-03-10 04:55:00	366	-1	2020-03-10 05:46:00	384	-1
2020-03-10 04:56:00	372	-1	2020-03-10 05:47:00	384	-1
2020-03-10 04:57:00	375	-1	2020-03-10 05:48:00	384	-1
2020-03-10 04:59:00	380	-1	2020-03-10 05:49:00	384	-1
2020-03-10 05:27:00	384	-1	2020-03-10 05:50:00	384	-1
2020-03-10 05:28:00	384	-1	2020-03-10 05:51:00	384	-1
2020-03-10 05:29:00	384	-1	2020-03-10 05:52:00	384	-1
2020-03-10 05:30:00	384	-1	2020-03-10 05:53:00	384	-1
			2020-03-10 05:54:00	384	-1

Рисунок 73 – Аномалии методом LOF

Теоретическое обоснование метода представлено в работе [55].

4.9.3.2 DBSCAN

DBSCAN – плотностный алгоритм кластеризации, выделяющий «внутренние», «пограничные» и «шумовые» точки в данных (точки, не вошедшие ни в один кластер). В основе алгоритма лежит идея о том, что плотность точек внутри каждого кластера существенно выше плотности точек снаружи, в том числе выше плотности шумовых областей.

Алгоритм DBSCAN выделяет в анализируемом временном ряде 4 кластера (таблица 10).

Таблица 10 – DBSCAN

Номер кластера	Количество наблюдений
----------------	-----------------------

0	1782
1	557
2	40
3	6

Соответственно «существенные» аномалии представлены в кластере 2 и 3 (рисунок 74.):

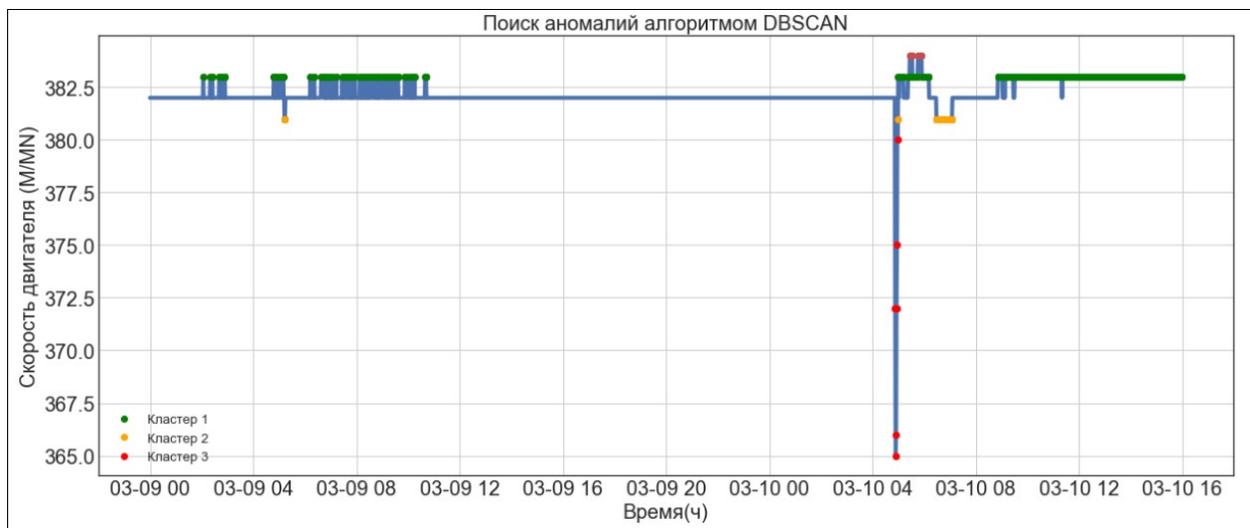


Рисунок 74 – Поиск аномалий алгоритмом DBSCAN

Теоретическое обоснование метода представлено в работе [56].

4.9.4 Методы машинного обучения

4.9.4.1 One Class SVM

Метод опорных векторов (support vector machine, SVM) (рисунок 75.) – один из методов обучения, применяемый для решения задач классификации и регрессии. Метод заключается в том, что строится гиперплоскость, которая разделяет объекты оптимальным образом. Поэтому необходимо настроить веса алгоритма для линейного порогового классификатора:

$$a(x) = \text{sign} \left(\sum_{j=1}^n w_j x^j - w_0 \right), \quad (26)$$

где w_j – веса алгоритма.

Их необходимо подобрать таким образом, чтобы объекты классов лежали, как можно дальше от разделяющей гиперплоскости – то есть для максимизации зазора.

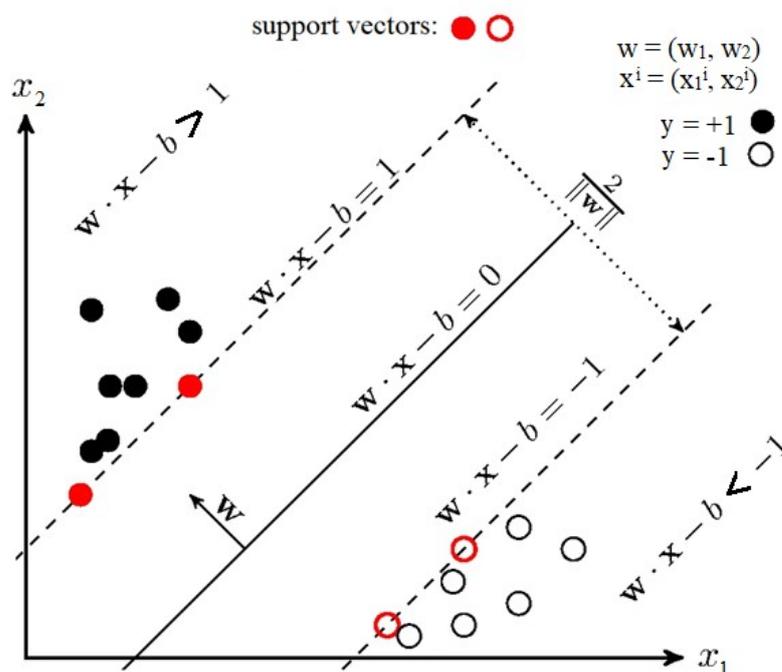


Рисунок 75 – SVM

SVM используется не только для линейно-разделимых выборок. Осуществляется это с помощью «ядра» (kernel). Ядро – это функция, способная преобразовать признаковое пространство, без преобразования признаков для того, чтобы классы линейно неразделимые в одном пространстве, разделить в другом.

One Class SVM – это одна из форма SVM, но для обучения достаточно иметь один класс. Алгоритмом строится граница по одну сторону которого схожие наблюдения, а по другую – аномалии.

Применительно к анализируемому временному ряду выделено 61-но аномальное наблюдение из 2400, их распределение по ряду приведено на рисунке 76:

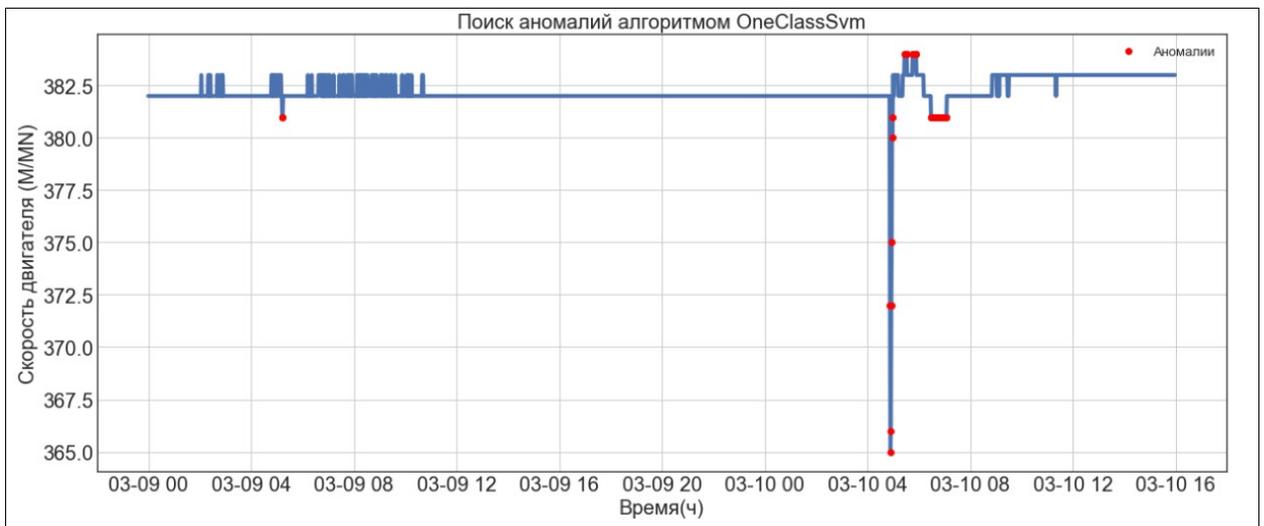


Рисунок 76 – Поиск аномалий алгоритмом OneClassSVM

Теоретическое обоснование метода представлено в работе [57].

4.9.4.2 Isolation Forest

Изолирующий лес (Isolation Forest) основан на принципе МонтеКарло: проводится случайное разбиение пространства признаков, такое что в среднем изолированные точки отсекаются от нормальных, кластеризованных данных. Окончательный результат усредняется по нескольким запускам стохастического алгоритма.

Алгоритм работает в контексте выявления аномалий с вычислением количества разделений дерева, требуемых для изоляции одного элемента выборки. Критерием останова является тождественное совпадение всех объектов в узле. Аномальные точки оказываются в листьях с низкой глубиной, так как «длина пути» рекурсивных разделений от корня дерева должна быть короче для аномальных наблюдений, чем для нормальных. Коэффициент контаминации определяет пороговое значение для наблюдений, которые должны считаться аномальными.

Применительно к анализируемому временному ряду модель Isolation Forest (рисунок 77.) находит определенное количество аномалий в зависимости от коэффициента контаминации. Чем ниже коэффициент, тем меньше аномалий обнаруживает модель.

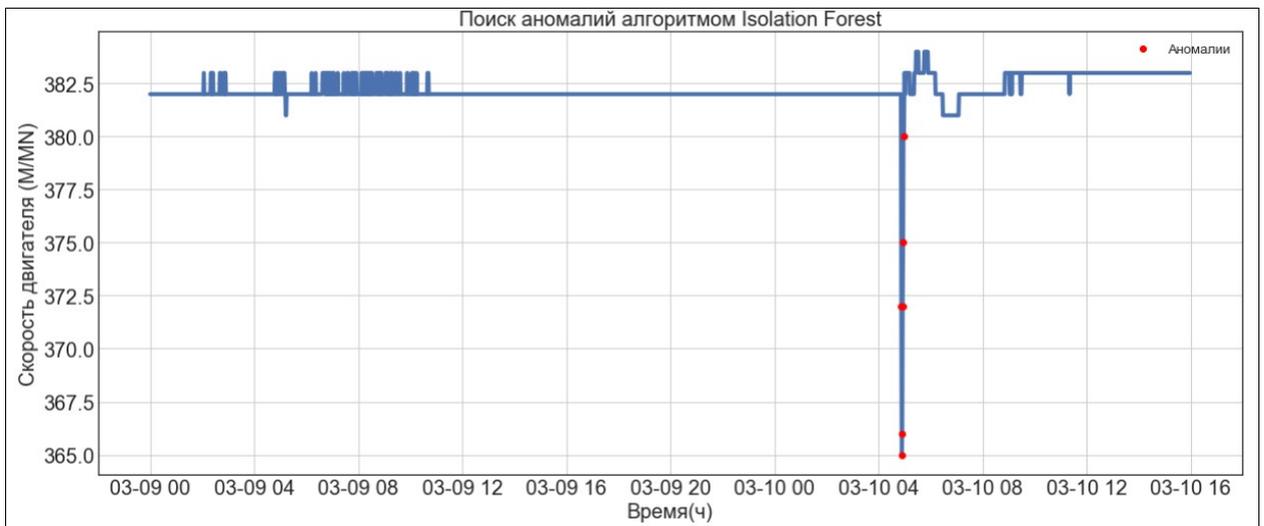


Рисунок 77 – Поиск аномалий алгоритмом Isolation Forest

Теоретическое обоснование метода представлено в работе [58].

4.9.5 Модельные методы

Модельный метод заключается в том, что строится модель, которая описывает данные – наблюдения, сильно отклоняющиеся от модели, есть аномалии.

Для анализируемого временного ряда используются модели Хольт-Уинтерса (см.4.8.2.3.) и линейной регрессии (см.4.8.5). Строятся доверительные интервалы на основе оценки дисперсии ошибки на один шаг вперед. Наблюдения, выходящие за доверительный интервал, считаются аномалиями.

Найденные аномалии модельными методами представлены на рисунках 78, 79:

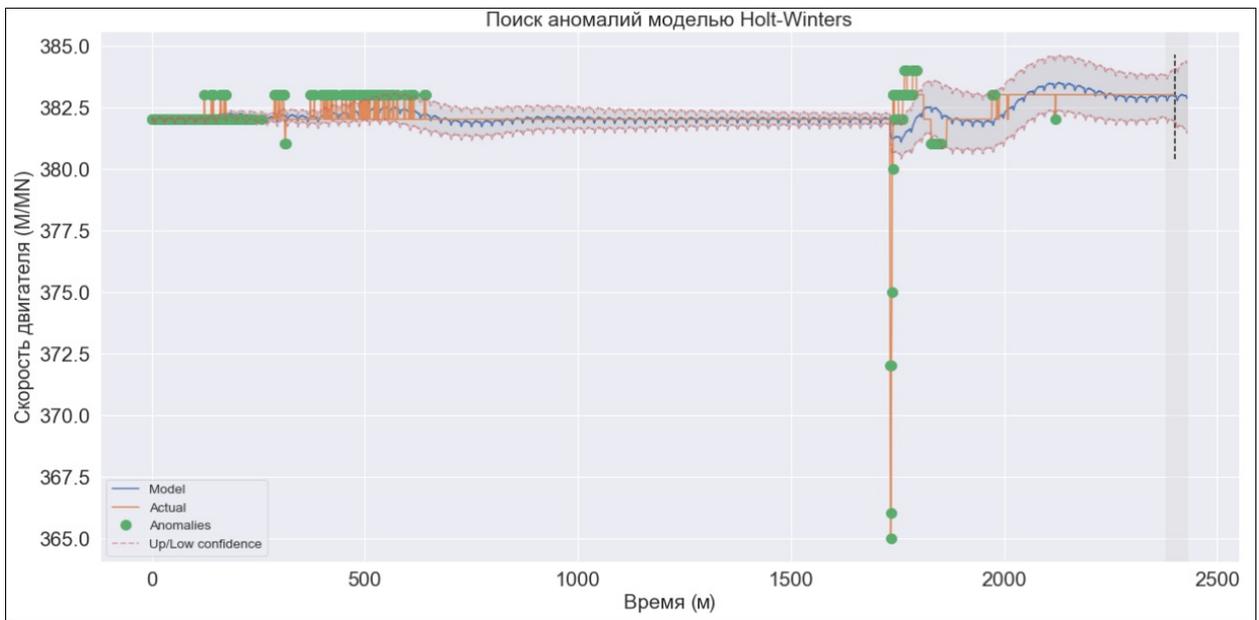


Рисунок 78 – Поиск аномалий моделью Холт-Уинтерса

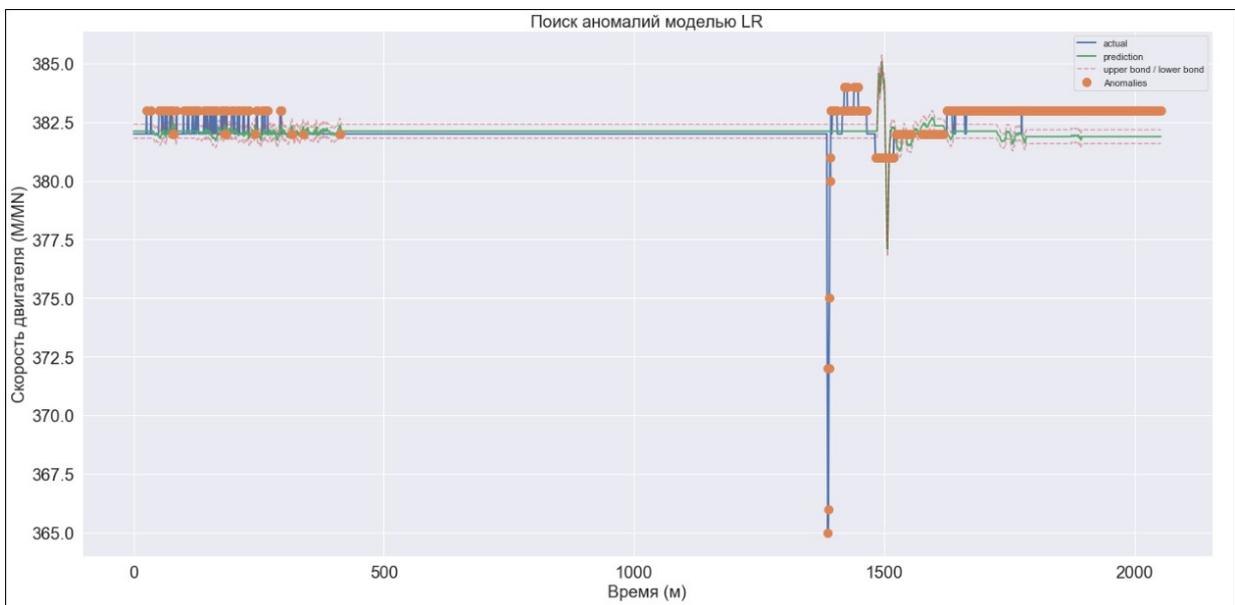


Рисунок 79 – Поиск аномалий моделью линейной регрессии

Теоретическое обоснование метода представлено в работе [51].

4.9.6 Методы на основе экспертных знаний

Метод основан на экспертных знаниях технологов, инженеров АСУ, проектировщиков производства. Рассчитываются пороговые значения для каждого из процессов/объектов технологической линии. Разрабатывается алгоритм работы линии при срабатывании какого-либо из порогов. Нахождение вне рамок диапазона пороговых значений соответствует

аномальному поведению. Простейшей модификацией, позволяющей снизить количество ложных срабатываний, является добавление счетчика, который накапливает события «выпадения» наблюдаемых параметров из диапазона. При превышении счетчиком определенного значения фиксируется факт наличия аномалии. Алгоритмы реализует ПЛК.

Общий вид продукционного правила следующий: *If condition THEN action*. В посылке данного правила содержится условие, необходимое для срабатывания алгоритма действий при пороговом значении. Когда все условия в левой части правила удовлетворены, выполняются действия, указанные в сукцеденте.

4.9.7 Сравнение методов детектирования

В разделах выше рассмотрены основные методы детектирования аномалий. Найдены аномалии анализируемого временного ряда.

В таблице 11 приведены достоинства и недостатки каждого из методов.

Таблица 11 – Сравнение методов поиска аномалий

Методы	Достоинства	Недостатки
Статистические методы	Интерпретируемость методов; простота использования; универсальность; не требуется априорной информации о причинах возникновения аномалии; отсутствие этапа предварительного обучения.	Подходит только для поиска выбросов, а аномалии характеризуются не только экстремальными значениями.

Продолжение таблицы 11 – Сравнение методов поиска аномалий

Методы	Достоинства	Недостатки
Контрольные карты	Интерпретируемость и наглядность метода; простота построения для небольшого количества данных; мониторинг процесса; отсутствие этапа предварительного обучения.	Метод морально устарел (в алгоритме выбора карты один из шагов предполагает выборку больше 10 наблюдений или меньше, что не соответствует положению дел); максимальное количество наблюдений в одной выборке – 25; ошибки первого и второго рода.
Модели, основанные на плотности	Наглядность и интерпретируемость метода; масштабируемость; высокая скорость моделирования; возможность обнаружения постепенных, но значительных отклонений от нормального поведения.	Присутствие этапа обучения; выбор метрики расстояния; при неправильном подборе метрики расстояния обнаружение аномалий, которые не оказывают существенного влияния на технологический процесс.
Методы машинного обучения	Наглядность метода; масштабируемость; высокая скорость моделирования; распознаются аномалии различных видов; не требует задания метрики; гибкость к динамике процесса.	Присутствие этапа обучения; сложность интерпретации; при неправильном подборе гиперпараметров обнаружение аномалий, которые не оказывают существенного влияния на технологический процесс.

Продолжение таблицы 11 – Сравнение методов поиска аномалий

Методы	Достоинства	Недостатки
Модельные методы	Наглядность метода; масштабируемость; не требуется знания о виде аномалий;	Присутствие этапа обучения; сложность интерпретации; выбор порога обнаружения аномалий;
Методы на основе экспертных знаний	Наглядность метода; простота реализации; отсутствие этапа обучения; простота интерпретации.	Необходимость высоких знаний экспертов при задании порогов; отсутствие адаптации; возможность пропуска аномалий.

Программный код одного из методов поиска аномалий приведен в приложении В.

4.10 Интеграция предиктивной модели в технологический процесс производства БОПП-пленки

4.10.1 Имитационная модель ТДО

Из-за отсутствия доступа к реальному технологическому процессу – разрабатывается имитационная модель части технологической линии, которая имитирует работу двигателей входа и выхода ТДО. Имитационная модель необходима для проверки работоспособности предиктивной модели; разработки архитектуры взаимодействия модели, среднего и верхнего уровня АСУ ТП; сбора информации в базу данных реального времени и онлайн мониторинг этой информации.

Для реализации имитационной модели используется пакет Matlab – Simulink. Simulink – среда динамического междисциплинарного моделирования сложных технических систем и основной инструмент для модельно-ориентированного проектирования.

На вход имитационной модели поступают данные (используемые ранее для разработки прогнозной модели и модели поиска аномалий), полученные с Vijeo Citect. На выходе эти же наблюдения, расположенные во времени с периодичностью в одну секунду.

Структурная схема имитационной модели в среде Simulink представлено на рисунке 80:

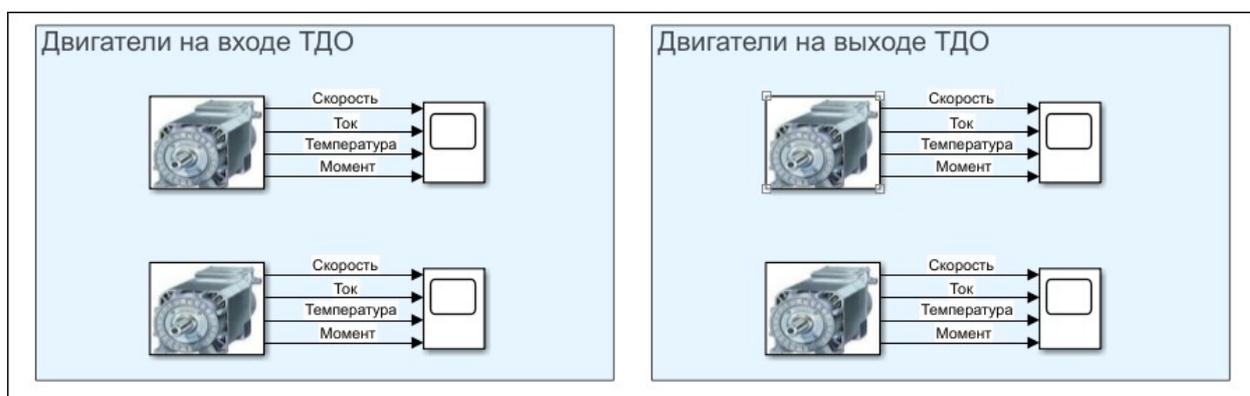


Рисунок 80 – Структурная схема имитационной модели
Выход двигателей представлен на рисунке 81:

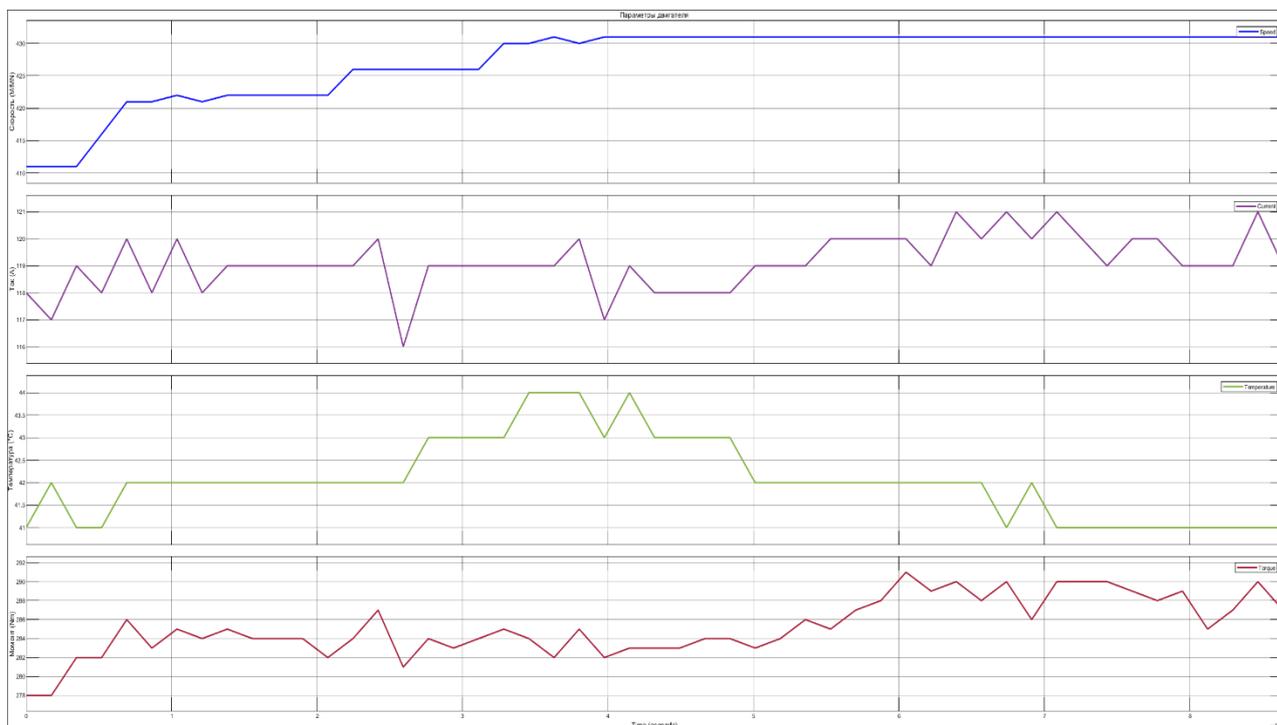


Рисунок 81 – Выход двигателей

4.10.2 SCADA-пакет

Автоматизированная система управления технологической линии производства БОПП-пленки реализуется на базе ПЛК Siemens S7-1500. Следовательно, в качестве платформы для разработки SCADA и алгоритмов выбирается TIA Portal [59].

TIA Portal (Totally Integrated Automation Portal) — интегрированная среда разработки программного обеспечения систем автоматизации технологических процессов от уровня приводов и контроллеров до уровня человеко-машинного интерфейса. Является воплощением концепции комплексной автоматизации (англ. Totally Integrated Automation) и эволюционным развитием семейства систем автоматизации Simatic компании Siemens AG.

TIA Portal содержит инструментальные средства для:

1. Конфигурирования и настройки параметров аппаратуры.
2. Конфигурирования систем промышленной связи.
3. Программирования контроллеров.

4. Конфигурирования панелей операторов SIMATIC.
5. Тестирования, выполнения пуско-наладочных работ и обслуживания готовой системы.

4.10.3 OPC-сервер

Для обеспечения корректной передачи данных в реальном времени между средой моделирования Simulink и TIA Portal используется программный пакет Process Simulator 2. Для передачи данных между TIA Portal и Python используется протокол OPC UA.

Платформа Process Simulator предназначена для тестирования систем автоматизации, начиная с имитации входных сигналов, и заканчивая построением полного симулятора технологического оборудования. Plug-in архитектура с открытым API позволяет любому программисту добавить недостающий функционал и пользоваться чужими наработками. Проект сохраняется в виде XML документа, а встроенный OPC сервер открывает доступ ко всем данным для любого внешнего клиента [60].

Протокол OPC UA – коммуникационный стандарт, определяющий передачу данных в промышленных сетях и взаимодействие устройств в них. OPC UA отличается от классических OPC тем, что полностью кроссплатформенный, простота удаленного подключения, возможность шифрования и аутентификации, а также унификация данных.

Siemens S7-1500 в базовой комплектации имеет протокол OPC UA, поэтому ПЛК может выступать и в качестве сервера, и в качестве клиента. На Python разработана библиотека `opc ua` [61], что позволяет реализовать получение данных с сервера Siemens S7-1500 и использовать их для предиктивной модели.

4.10.4 База данных

На производстве БОПП-пленки используется SCADA Vijeo Citect, которая хранит исторические данные за 3 месяца. Из-за малого количества исторических данных сужается спектр возможных внедрённых цифровых технологий, отсутствует ретроспективный анализ технологического процесса и т.д.. Возникает потребность в базе данных.

Для хранения данных технологического процесса лучше всего подходит база данных временных рядов. Базы данных временных рядов предоставляют пользователям сервис по хранению, удалению, обновлению и модификации временных рядов. Реализуется поддержка некоторых вычислений над временными рядами. Существуют инструменты для фильтрации по произвольным образцам, при этом значения одного ряда могут являться фильтром для другого. Преимущества баз данных временных рядов перед реляционными, нереляционными распределёнными, циклическими базами данных описаны в работе [62].

Среди баз данных временных рядов была выбрана InfluxDB, исходя из статей [62], [63]. InfluxDB имеет SQL подобный язык запросов – InfluxQL, HTTP(S) API для записи и выборки данных, встроенную поддержку других популярных в мониторинге протоколов передачи данных, возможность сохранять миллиарды точек измерений, слияние временных рядов друг с другом.

4.10.5 Онлайн-мониторинг технологического процесса

При долговременном отслеживании параметров технологического процесса (больше, чем 3 месяца) возникает потребность не только в базе данных, но и в системе онлайн-мониторинга базы данных. Для этого выбрана платформа Grafana.

Grafana — это платформа с открытым исходным кодом для визуализации, мониторинга и анализа данных, часто используемая в

сочетании с базами данных временных рядов, в том числе, с InfluxDB [64].

Для подключенной к ней базе данных предоставляет:

1. Менеджер уведомлений (для отслеживания отклонений технологического процесса).
2. Форму для запросов к данным.
3. Построение счетчиков, шкал, гистограмм.
4. расчет базовых метрик (среднее арифметическое, максимальное, минимальное значение и т.д.).
5. Дашборды для визуализации временных рядов и сопутствующих им метрик.

4.10.6 Реализация взаимодействия технологического процесса и предиктивной модели

Технологический процесс представлен имитационной моделью (см. 4.10.1) (для примера используется один двигатель выхода ТДО). Связь с Process Simulator устанавливается с помощью OPC Toolbox. OPC Toolbox предоставляет доступ к оперативным и архивным данным OPC напрямую из MATLAB и Simulink. Конфигурация соединения осуществляется блоком OPC Configuration. Блоки OPC Read и OPC Write извлекают и передают на сервер и обратно. Настройки имитационной модели в Simulink представлены на рисунках 82 и 83:

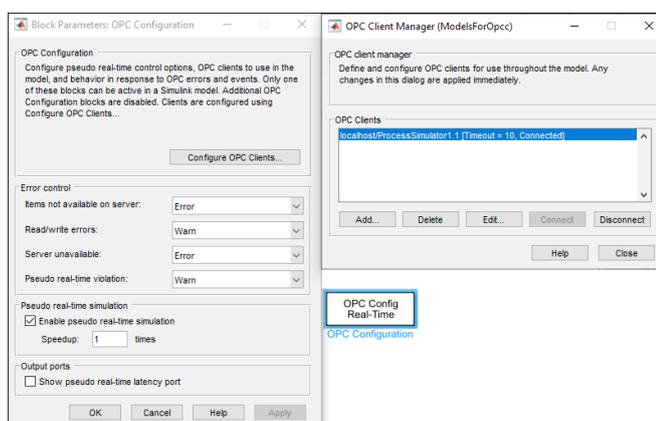


Рисунок 82 – Настройка блока OPC Configuration

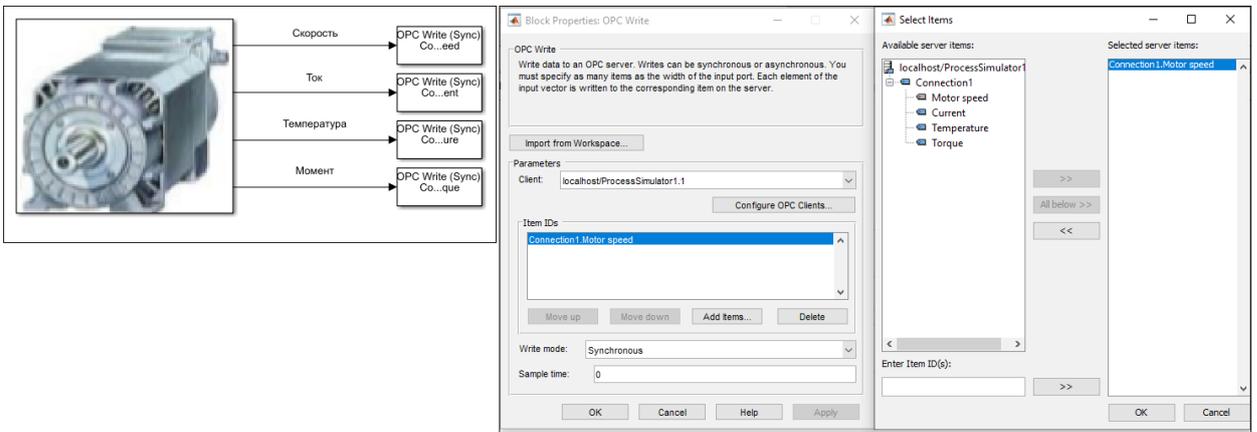


Рисунок 83 – Настройка передачи информации в Process Simulator и блока OPC Write

Данные с Simulink поступают в Process Simulator (рисунок 84.), который, в свою очередь, передает их в TIA Portal (рисунок 85.). Мониторинг значений в TIA Portal реализуется с помощью Force Table.

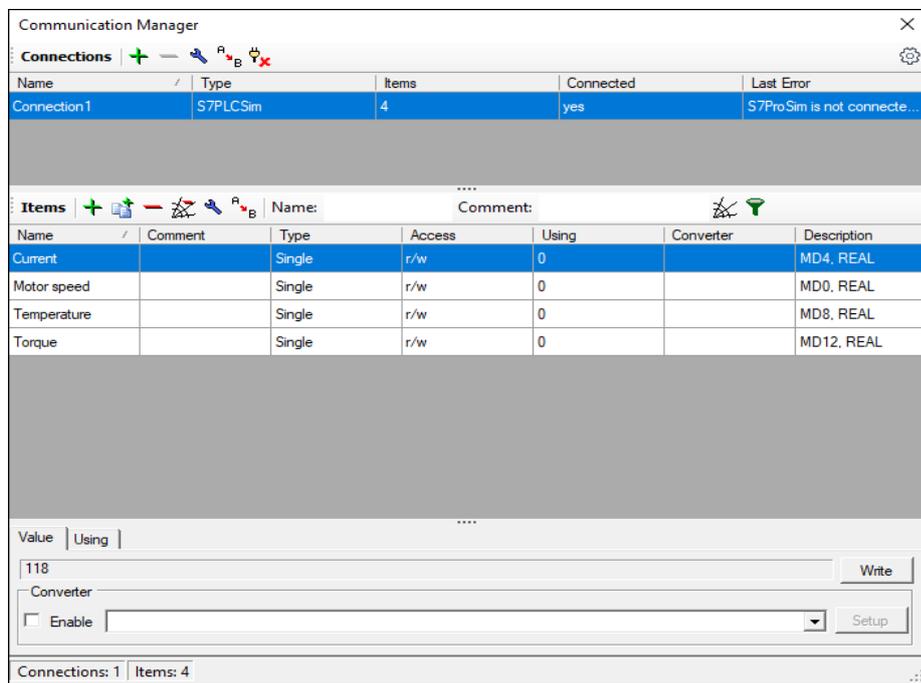


Рисунок 84 – Process Simulator

i	Name	Address	Display format	Monitor value
1	"Motor speed"	%MD0	Floating-point nu...	411.0
2	"Current"	%MD4	Floating-point nu...	118.0
3	"Temperature"	%MD8	Floating-point nu...	41.0
4	"Torque"	%MD12	Floating-point nu...	282.0

Рисунок 85 – Онлайн мониторинг значений имитационной модели в TIA Portal

Для отправки данных с TIA Portal необходимо настроить виртуальное ethernet соединение с контроллером Siemens S7-1500, который, в свою очередь, должен выступать в роли OPC-сервера (рисунок 86., рисунок 87.):

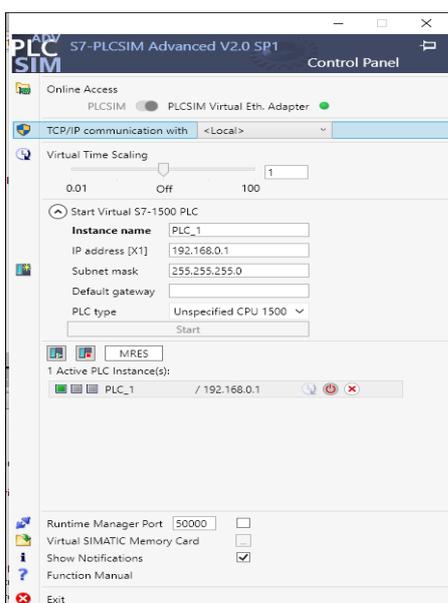


Рисунок 86 – Настройка виртуального ethernet адаптера

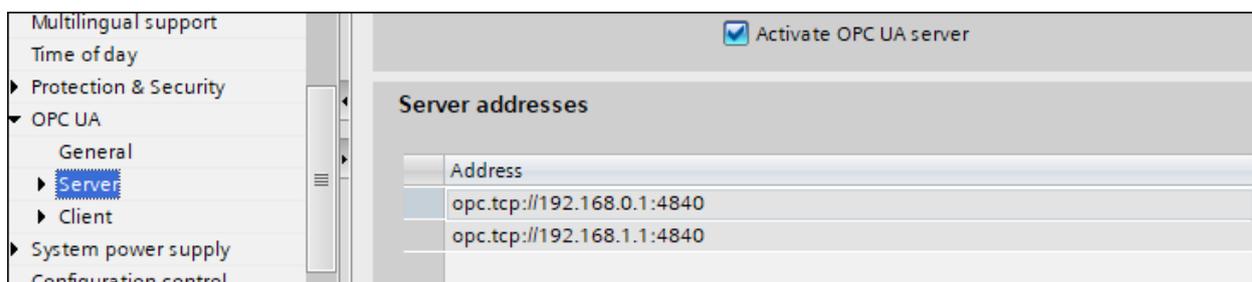


Рисунок 87 – Активация OPC-сервера

После активации контроллера в роли OPC-сервера к нему подключается UaExpert (рисунок 88.):

#	Server	Node Id	Display Name	Value	Datatype	Source Timestamp	Server Timestamp	Statuscode
1	Server	NS3 String "Motor speed"	Motor speed	411	Float	20:01:36.298	20:01:36.298	Good
2	Server	NS3 String "Current"	Current	118	Float	20:01:40.298	20:01:40.298	Good
3	Server	NS3 String "Temperature"	Temperature	41	Float	20:01:45.298	20:01:45.298	Good
4	Server	NS3 String "Torque"	Torque	282	Float	20:01:47.298	20:01:47.298	Good

Рисунок 88 – Параметры двигателя, переданные по протоколу OPC UA в UaExpert

UaExpert – полнофункциональный клиент OPC UA, с помощью которого можно передать данные в Python, выступающего в роли клиента по отношению к UaExpert. Для этого используется библиотека Python opc-ua (исходный код представлен в приложении Г).

Дальнейшая передача информации в базу данных InfluxDB производится с помощью библиотеки influxdb [65], предварительно приведя в верный формат дату и время, которые определены в RFC3339. Добавление новой точки измерения выглядит следующим образом:

Таблица 12 – Формат записи новой точки в InfluxDB

Measurement	Tags	Time	Fields
-------------	------	------	--------

Информация, переданная в InfluxDB, представлена на рисунке 89:

```

> show databases
name: databases
name
----
_internal
Engine
Current (Actual)
Current (Forecast)
Motor speed (Actual)
Motor speed (Forecast)
Temperature (Actual)
Temperature (Forecast)
Torque (Actual)
Torque (Forecast)

```

Рисунок 89 – База данных InfluxDB и записанные в неё параметры двигателя

Онлайн-мониторинг параметров двигателей и предсказанных значений реализуется системой Grafana (рисунок 90.):



Рисунок 90 – Онлайн-мониторинг параметров двигателя, предсказанных значений и базовых метрик

Схема информационных потоков для представлена в приложении Д.

5 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

Научно-исследовательская работа направлена на разработку программного модуля, осуществляющего прогнозную аналитику технологического процесса и детектирования аномалий в нем. Технологический процесс представляет собой производство БОПП-пленки.

Целью данной главы является определение потенциала разрабатываемого модуля, планирование процесса управления исследованием, определение ресурсной, финансовой и экономической эффективности.

Достижение цели обеспечивается решением следу задач:

1. организация работ по разрабатываемому проекту;
2. планирование работ по проекту;
3. определение ресурсной, финансовой, бюджетной, социальной и экономической эффективности исследования.

5.1 Организация и планирование работ

Занятость каждого из участников при проведении каждого из этапов работ должна быть оптимально спланирована по срокам. На каждом этапе выполнения работ определяются исполнители и продолжительность каждого этапа. Календарный график реализации проекта – это результат планирования работ. Он представляет из себя наиболее наглядный и удобный способ организации проекта. Для его построения определяются даты начала и окончания работ, а также составляется перечень работ и соответствие работ своим исполнителям. Основные исполнители в проекте: научный руководитель (НР) и инженер (И).

Разделение выполнения дипломной работы на этапы представлены в таблице 13.

Таблица 13 – Перечень работ и продолжительность их выполнения

Этапы работы	Исполнитель и	Загрузка исполнителей
Определение темы ВКР	НР, И	НР – 30% И – 70%
Поиск и изучение нормативно-технической литературы	НР, И	НР – 10% И – 100%
Календарное планирование работ	НР, И	НР – 10% И – 100%
Выбор технологического процесса	И	И – 100%
Сбор данных с технологического процесса	И	И – 100%
Описание технологического процесса (изучение оборудования и процесса)	И	И – 100%
Анализ методов прогнозирования и детектирования аномалий	И	И – 100%
Обработка данных технологической линии	И	И – 100%
Математическое описание используемых моделей	И	И – 100%
Построение моделей (выбор параметров, обучение на наборе данных)	И	И – 100%
Тестирование моделей (визуализация результатов)	И	И – 100%
Сравнение результатов работы моделей	И	И – 100%
Написание раздела «финансовый менеджмент, ресурсоэффективность и ресурсосбережение»	И	И – 100%
Написание раздела «социальной	И	И – 100%
Составление пояснительной записки	И	НР – 80 И – 20%
Проверка работы с руководителем	НР, И	НР – 20% И – 80%
Подготовка презентации дипломного проекта	И	И – 100%

5.1.1 Продолжительность этапов работ

Для определения ожидаемого (среднего) значения трудоемкости используется следующая формула:

$$t_{ож} = \frac{3*t_{mini} + 2*t_{maxi}}{5}, \quad (27)$$

где t_{min} – минимальная трудоемкость i -ой работы, чел/дн.;

t_{max} – максимальная трудоемкость i -ой работы, чел/дн.

Для построения линейного графика необходимо рассчитать длительность этапов в рабочих днях, а затем перевести ее в календарные дни. Расчет продолжительности выполнения каждого этапа в рабочих днях (T_{PD}) ведется по формуле:

$$T_{PD} = \frac{t_{ож}}{K_{ВН}} \cdot K_{Д} T_{PD} = \frac{t_{ож}}{K_{ВН}} \cdot K_{Д}, \quad (28)$$

где $t_{ож}$ – продолжительность работы, дн.;

$K_{ВН}$ – коэффициент выполнения работ, учитывающий влияние внешних факторов на соблюдение предварительно определенных длительностей ($K_{ВН} = 1$);

$K_{Д}$ - коэффициент, учитывающий дополнительное время на компенсацию непредвиденных задержек и согласование работ ($K_{Д} = 1,2$).

Расчет продолжительности этапа в календарных днях ведется по формуле:

$$T_{КД} = T_{PD} \cdot T_{К}, \quad (29)$$

где $T_{КД}$ – продолжительность выполнения этапа в календарных днях;

$T_{К}$ – коэффициент календарности, позволяющий перейти от длительности работ в рабочих днях к их аналогам в календарных днях, и рассчитываемый по формуле:

$$T_{К} = \frac{T_{КАЛ}}{T_{КАЛ} - T_{ВД} - T_{ПД}}, \quad (30)$$

где $T_{КАЛ}$ – календарные дни;

$T_{ВД}$ – выходные дни;

$T_{ПД}$ – праздничные дни.

Рассчитаем коэффициент календарности для пятидневной рабочей недели из расчета 145 нерабочих дней на 2021 год:

$$T_{К} = \frac{T_{кал}}{T_{кал} - T_{вых} - T_{пр}} = \frac{365}{365 - 118} = 1,47 \quad (31)$$

В таблице 14 приведены продолжительности этапов работ и их трудоемкости по исполнителям, занятым на каждом этапе. Стоит отметить,

что величины трудоемкости этапов по исполнителям $T_{кд}$ позволяют построить линейный график осуществления проекта, представленный на рисунке 1.

Таблица 14 – Трудозатраты на выполнение проекта

Этап	Исполнитель и	Продолжительность работ, дни			Трудоемкость работ по исполнителям чел.- дн.			
					$T_{рд}$		$T_{кд}$	
		t_{min}	t_{max}	$t_{ож}$	НР	И	НР	И
Определение темы ВКР	НР, И	2	4	2,8	1,4	4,5	0,3	1
Поиск и изучение нормативно-технической литературы	НР, И	3	5	3,8	0,6	6,8	0,1	1,1
Календарное планирование работ	НР, И	6	8	6,8	1,1	12,1	0,1	1,1
Выбор технологического процесса	И	2	3	2,4	0,4	3,9	0,1	1
Сбор данных с технологического процесса	И	6	8	6,8	0,0	11,0	0	1
Описание технологического процесса	И	5	7	5,8	0,0	9,4	0	1
Анализ методов прогнозирования и детектирования аномалий	И	5	6	5,4	0,0	8,7	0	1

Этап	Исполнитель	Продолжительность работ, дни			Трудоемкость работ по исполнителям чел.-дн.			
					$T_{рД}$		$T_{кД}$	
		t_{min}	t_{max}	$t_{ож}$	НР	И	НР	И
Обработка данных технологической линии	НР, И	6	8	6,8	0,0	14,3	0	1,3
Математическое описание используемых моделей	И	6	8	6,8	0,0	11,0	0	1
Построение моделей (выбор параметров, обучение на наборе данных)	И	8	9	8,4	0,0	13,6	0	1
Тестирование моделей (визуализация результатов)	И	4	5	4,4	0,0	7,1	0	1
Сравнение результатов работы моделей	И	10	12	10,8	0,0	24,4	0	1,4
Написание раздела «финансовый менеджмент, ресурсоэффективность и	И	3	5	3,8	0,0	6,1	0	1

Этап	Исполнитель	Продолжительность работ, дни			Трудоемкость работ по исполнителям чел.-дн.			
					$T_{РД}$		$T_{КД}$	
		t_{min}	t_{max}	$t_{ож}$	НР	И	НР	И
ресурсосбережение»								
Написание раздела «социальной	НР, И	2	3	2,4	0,0	3,9	0	1
Составление пояснительной записки	НР, И	2	3	2,4	3,9	0,4	1	0,1
Проверка работы с руководителем	И	5	7	5,8	2,8	9,4	0,3	1
Подготовка презентации дипломного проекта		2	4	2,8	0,0	4,5	0	1
Итого				88,2	6,908	102,806	10,1547	151,1248

На основе календарного плана проекта построена диаграмма Ганта (рисунок 91), которая наглядно показывает следование выполнения этапов дипломного проектирования, исходя из отведенных сроков. Диаграмма Ганта – это тип столбчатых диаграмм, который используется для иллюстрации календарного плана проекта, на котором работы по теме представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения данных работ.

Этап	НР	И	Февраль			Март			Апрель			Май			Июнь		
			10	20	30	40	50	60	70	80	90	100	110	120	130	140	150
Определение темы ВКР	1,36	4,53	■														
Поиск и изучение нормативно-технической литературы	0,61	6,76	■	■													
Календарное планирование работ	1,10	12,10		■	■												
Выбор технологического процесса	0,39	3,88			■												
Сбор данных с технологического процесса	0,00	11,00				■	■										
Описание технологического процесса	0,00	9,38					■	■									
Анализ методов прогнозирования и детектирования аномалий	0,00	8,73						■	■								
Обработка данных технологической линии	0,00	14,29							■	■							
Математическое описание используемых моделей	0,00	11,00								■	■						
Построение моделей (выбор параметров, обучение на наборе данных)	0,00	13,58									■	■					
Тестирование моделей (визуализация результатов)	0,00	7,11										■					
Сравнение результатов работы моделей	0,00	24,45											■	■			
Написание раздела «финансовый менеджмент, ресурсоэффективность и ресурсосбережение»	0,00	6,14												■			
Написание раздела «социальной	0,00	3,88													■		
Составление пояснительной записки	3,88	0,39														■	
Проверка работы с руководителем	2,81	9,38														■	■
Подготовка презентации дипломного проекта	0,00	4,53															■

Рисунок 91 – Линейный график осуществления проекта

5.2 Расчет сметы затрат на выполнение проекта

При планировании бюджета исследования должно быть обеспечено полное и достоверное отражение всех видов планируемых расходов, необходимых для его выполнения.

В состав затрат на создание проекта включается величина всех расходов, необходимых для реализации комплекса работ, составляющих содержание данной разработки. Расчет сметной стоимости ее выполнения производится по следующим статьям затрат:

1. материалы и покупные изделия;
2. заработная плата;
3. социальный налог;
4. расходы на электроэнергию (без освещения);
5. амортизационные отчисления;
6. командировочные расходы;
7. оплата услуг связи;
8. арендная плата за пользование имуществом;
9. прочие услуги (сторонних организаций);
10. прочие (накладные) расходы.

5.2.1 Затраты на материалы и покупные изделия

К данной статье расходов относится стоимость материалов, покупных изделий, полуфабрикатов и других материальных ценностей, расходуемых непосредственно в процессе выполнения работ над объектом проектирования. Кроме того, статья включает так называемые транспортно-заготовительные расходы.

Для выполнения расчетов расходы принимаются как 15 % от отпускной цены закупаемых материалов. Затраты на материалы сведены в таблице 15.

Таблица 15 – Расчет затрат на материала

Наименование материалов	Цена за ед., руб.	Кол-во	Сумма, руб.
PyCharm	13 957	1 экз.	13 957
Microsoft Office 2016 (лицензионное ПО)	3400	1 шт.	3400
Лицензионное ПО MatLaB	5600	1 экз.	5600
Бумага для принтера формата А4	250	1 уп.	250
Картридж для принтера	1800	1 шт.	1800
Итого			25007

Расходы на материалы с учетом ТЗР составляют:

$$С_{\text{мат}} = 25007 * 1,15 = 28758 \text{ руб.} \quad (32)$$

5.2.2 Затраты на заработную плату

Данная статья расходов включает заработную плату работников, непосредственно занятых выполнением НИИ (в том числе премии, доплаты и т.д)

Среднедневная тарифная заработная плата ($ЗП_{\text{дн-т}}$) рассчитывается по формуле:

$$ЗП_{\text{дн-т}} = МО/20,6 \quad ($$

$$\text{Дневная з/плата} = \frac{\text{Месячный оклад}}{25,17 \text{ дней}}$$

учитывающей, что в году 247 рабочих дней (на 2021 год) и, следовательно, в месяце в среднем 20,6 рабочих дня (при пятидневной рабочей неделе).

Расчеты затрат на полную заработную плату приведены в таблице 4. Затраты времени по каждому исполнителю в рабочих днях с округлением до целого взяты из таблицы 2.

Для учета в ее составе премий, дополнительной зарплаты и районной надбавки используется следующие коэффициенты (с учетом 5-ти дневной рабочей недели): $K_{\text{пр}} = 1,1$; $K_{\text{доп.ЗП}} = 1,113$ (используется для 6-ти дневной рабочей недели); $K_{\text{р}} = 1,3$. Таким образом, для перехода от тарифной (базовой) суммы заработка исполнителя, связанной с участием в проекте, к соответствующему полному заработку (зарплатной части сметы) необходимо первую умножить на интегральный коэффициент:

$$K_{и}=1,1*1,118*1,3 = 1,6. \quad (34)$$

Таблица 16 – Затраты на заработную плату

Исполнитель	Оклад, руб./мес.	Среднедневная ставка, руб./раб.день	Затраты времени, раб.дни	Коэффициент	Фонд з/п., руб.
НР	33 664	1634	14	1,6	36601,6
И	13 500	655	82	1,6	85936
Итого					122537,6

5.2.3 Затраты на социальный налог

Затраты на единый социальный налог (ЕСН), включающий в себя отчисления в пенсионный фонд, на социальное и медицинское страхование, составляют 30% от полной заработной платы по проекту, т.е. $C_{соц} = K_{соц} \cdot C_{осн}$
 $C_{соц} = C_{зп} * 0,3$. Тогда расходы на социальный налог составляют:

$$C_{соц} = 122537,6 * 0,3 = 36761,28 \text{ руб.} \quad (35)$$

5.2.4 Затраты на электроэнергию

Данный вид расходов включает в себя затраты на электроэнергию, потраченную при выполнении проекта на работу используемого оборудования, рассчитываемые по формуле:

$$C_{эл.об.} = P_{об} \cdot t_{об} \cdot Ц_{э}, \quad (36)$$

где $P_{об}$ – мощность, потребляемая оборудованием, кВт, $Ц_{э}$ – тариф на 1 кВт·час, $t_{об}$ – время работы оборудования, час.

Для ТПУ $Ц_{э} = 6,59$ руб./квт·час (с НДС).

Время работы оборудования вычисляется на основе затраченного времени (см. таблицу 2) для инженера ($T_{рд}$) из расчета 8-ми часового рабочего дня:

$$t_{об} = T_{рд} * K_t, \quad (37)$$

где $K_t \leq 1$ – коэффициент использования оборудования по времени, равный отношению времени его работы в процессе выполнения проекта к $T_{РД}$, примем данный коэффициент равным 0,8.

Мощность, потребляемая оборудованием, определяется по формуле:

$$P_{ОБ} = P_{ном.} * K_C, \quad (38)$$

где $P_{ном.}$ – номинальная мощность оборудования, кВт, $K_C \leq 1$ – коэффициент загрузки, зависящий от средней степени использования номинальной мощности. Для технологического оборудования малой мощности $K_C = 1$.

Затраты на электроэнергию для технологических целей приведен в таблице 17.

Таблица 17 – Затраты на электроэнергию технологическую

Наименование оборудования	Время работы оборудования $t_{об}$, час	Потребляемая мощность $P_{об}$, кВт	Затраты $\mathcal{E}_{об}$, руб.
Персональный компьютер	$81,4 * 8 * 0,8 = 520,96$	0,3	1029,93
Лазерный принтер	2	0,1	1,32
Итого			1031,25

5.2.5 Амортизационные расходы

В статье «Амортизационные отчисления» рассчитывается амортизация используемого оборудования за время выполнения проекта.

Используется формула:

$$\frac{C_{об} * t_{рф} * n}{F_d}, \quad (39)$$

где N_A – годовая норма амортизации единицы оборудования;

$C_{об}$ – балансовая стоимость единицы оборудования с учетом ТЗР;

F_d – действительный годовой фонд времени работы соответствующего оборудования, берется из фактического режима его использования в текущем календарном году;

$t_{рф}$ – фактическое время работы оборудования в ходе выполнения проекта;

n – число задействованных однотипных единиц оборудования.

Определение N_A содержится в постановлении правительства РФ «О классификации основных средств, включенных в амортизационные группы», согласно которому граничные значения сроков амортизации (полезного использования) оборудования (для ПК сроки составляют 2 ÷ 3 года). Для расчетов принимается значение 2 года. Значение величины N_A обратно пропорционально значению срока амортизации ($N_A = 0,5$). Для лазерного принтера принимается аналогичное значение.

Расчет C_{AM} для ПК выглядит следующим образом. Стоимость ПК 70000 руб., время использования 654,4 часа (см. таблицу 5), $N_A = 0,5$, $F_D = 220 * 8 = 1760$ часов (для пятидневной рабочей недели):

$$C_{AM} = \frac{0,5 * 70000 * 654,4 * 1}{1760} = 13013,63 \text{ руб.} \quad (40)$$

Значения параметров для принтера. Стоимость принтера 10000 руб., время использования 2 часа, $N_A = 0,5$, $F_D = 440$ часов:

$$C_{AM} = \frac{0,5 * 10000 * 2 * 1}{440} = 22,72 \text{ руб.} \quad (41)$$

Итого начислено амортизации 13036,35 руб.

5.2.6 Расходы, учитываемые непосредственно на основе платежных документов

Здесь учитывается оплата услуги интернет-связи за 5 месяцев (с 01.02 по 01.06) при ежемесячной плате 330 рублей. Таким образом, $C_{np1} = 330 * 5 = 1650$ руб.

Также к данному разделу расходов относится и подписка на платные интернет-издания, публикующие научные статьи. За аналогичный период времени $C_{np2} = 180 * 5 = 900$ руб.

Таким образом $C_{np} = C_{np1} + C_{np2} = 2350$ руб.

5.2.7 Прочие расходы

В статье «Прочие расходы» отражены расходы на выполнение проекта, которые не учтены в предыдущих статьях (принимаются равными 10 % от суммы всех вышеуказанных расходов):

$$C_{\text{проч.}} = (C_{\text{мат}} + C_{\text{зп}} + C_{\text{соц}} + C_{\text{эл.об.}} + C_{\text{ам}} + C_{\text{нп}}) \cdot 0,1 = \\ (18277 + 122537 + 36761,28 + 1031 + 13036 + 2350) \cdot 0,1 = 19399 \text{ руб.} \quad (42)$$

5.2.8 Общая себестоимость разработки

Проведя расчет по всем статьям сметы затрат на разработку, можно определить общую себестоимость проекта.

$$C_{\text{проч.}} = C_{\text{мат}} + C_{\text{зп}} + C_{\text{соц}} + C_{\text{эл.об.}} + C_{\text{ам}} + C_{\text{нп}} + C_{\text{проч.}} \\ = 193990 + 19399 = 213389 \text{ руб.} \quad (43)$$

5.2.9 Доход

Поскольку работа заключается в проведении исследования, результаты которого могут использоваться в различных предприятиях, условиях и т.д, то для получения дохода выручка с проекта должна превышать себестоимость (для расчета принимается 10 % превышение от полной себестоимости проекта).

Таким образом доход составляет 21338,9 руб.

5.2.10 Затраты на НДС

На 2021 год НДС в РФ составляет 20% от суммы затрат на разработку и дохода. Тогда затраты на НДС составляю

$$Ц_{\text{НДС}} = (213389 + 21338,9) \cdot 0,2 = 46945,58 \text{ руб.} \quad (44)$$

5.2.11 Цена разработки НИР

Цена равна сумме полной себестоимости, прибыли и НДС:

$$C_{\text{НИР(КР)}} = 213389 + 21338,9 + 46945,58 = 281673,48 \text{ руб.} \quad (45)$$

5.3 Оценка экономической эффективности проекта

На данный момент существует несколько методов обслуживания оборудования – аварийное, регламентное и предиктивное.

Аварийное обслуживание – обслуживание оборудования после выхода его из строя. Данный метод оправдан при обслуживании дешевого оборудования, не оказывающего решающего влияния на технологический процесс, либо имеющего резервирование.

Регламентное обслуживание осуществляется плановыми проверками согласно документации оборудования. План проверок рассчитывается с помощью статистического анализа компанией-изготовителем. Проблема возникает при работе оборудования на высоких нагрузках, что повышает шанс выхода из строя раньше плана проверки и соответственно приводит к нарушению технологического процесса.

Третий вид обслуживания – предиктивный. Когда процесс контролируется непрерывно по фактическому состоянию. Часто происходит так, что отклонения производственного процесса от обычного своего течения обнаруживается только по факту самого отклонения. Выявление таких отклонений (аномалий) происходит с помощью предсказательной аналитики, то есть строится модель предсказаний, имитирующая ход процесса и прогнозирующая его на несколько шагов вперед, после чего возникает подзадача – детектирование аномалий в прогнозируемой части процесса. Данный метод позволяет сформировать мероприятия технического обслуживания к моменту отклонения процесса, что уменьшает количество фактических неисправностей, простоев, затрат на обслуживание и т.д.

Данная работа посвящена применению методов машинного обучения для предиктивной аналитики технологического процесса производства БОПП-пленки с последующим детектированием аномалий.

Производство БОПП-пленки – это сложный технологический процесс, в котором некорректная работа какого-либо оборудования может привести к большому времени простоев, что, в свою очередь, приведет к значительным экономическим потерям. Поэтому предсказание процесса и выявление в нём аномалий существенно увеличит экономическую эффективность производства, так как уменьшит вероятность нарушения хода работы технологической линии, уменьшит потерю сырья при обрывах пленки и времени на ремонт оборудования.

6 Социальная ответственность

В разделе социальная ответственность рассматриваются вопросы обнаружения и анализа вредных и опасных факторов труда на рабочем месте инженера, минимизация негативных последствий проектируемой деятельности в соответствии с требованиями санитарных норм и правил, техники безопасности и пожарной безопасности.

Цель исследования – потенциальное уменьшение количества остановок производства БОПП-пленки и планирование мероприятий по обслуживанию оборудования за счет предиктивных моделей.

Предмет исследования – методы прогнозирования временного ряда и детектирования аномалий в нём.

Задачи исследования:

6. Сбор данных.
7. Подготовка данных.
8. Построение прогнозных моделей.
9. Детектирование аномалий.
10. Интеграция в технологический процесс.

Объект исследования – исторические данные (в форме временных рядов) о технологическом процессе: значения датчиков, исполнительных механизмов и другие параметры технологической линии.

Написание выпускной квалификационной работы должно было проходить на кафедре ОАР ИШИТР в рамках преддипломной практики, но в связи с коронавирусной инфекцией, работа производилась в условиях самоизоляции.

В рамках выполнения ВКР было сделано следующее:

1. подготовка к исследованию (выбор темы ВКР, изучение литературы, планирование работ, выбор технологического процесса);
2. подготовка данных (сбор, анализ, тесты);
3. построение прогнозной модели (анализ методов, математическое описание методов и их сравнение);

4. построение модели поиска аномалий (анализ методов, математическое описание методов и их сравнение);

5. интеграция разработанных моделей в имитационный технологический процесс.

6.3 Правовые и организационные вопросы обеспечения безопасности

6.3.1 Особенности трудового законодательства

Режим рабочего времени устанавливается в соответствии с «ТК РФ Статья 91. Понятие рабочего времени. Нормальная продолжительность рабочего времени» [10]. Нормальная продолжительность рабочего времени не может превышать 40 часов в неделю. При 5-ти дневной рабочей неделе 8-часовой рабочей смене и работе на ПК регламентированные перерывы следует устанавливать через 2 часа от начала смены и через 2 часа после обеденного перерыва продолжительностью 15 минут каждый.

Соблюдение защиты персональных данных устанавливается в соответствии с «ТК РФ Статья 86. Общие требования при обработке персональных данных работника и гарантии их защиты».

Оплата и нормирование труда производится согласно «ТК РФ Раздел VI. Оплата и нормирование труда». Оклад установлен в размере 30 000 тысяч рублей. Выполнение работ производится, исходя из плана работ, установленными научным руководителем.

Оплата работников, занятых на работах с вредными или опасными условиями труда выполняется в соответствии ТК РФ Статья 147. Так как условия труда не подходят под понятие опасные, следовательно выплат и компенсаций не производится.

6.3.1 Особенности трудового законодательства

Трудовой кодекс РФ и СанПиН 2.2.2/2.4.1340-03 определяют основные нормы безопасности трудовой деятельности и к рабочему месту сотрудника соответственно.

Работа с компьютером характеризуется умственным напряжением и высокой напряженностью зрительной работы, поэтому большое значение имеет расположение элементов рабочего места для поддержания оптимальной рабочей позы человека. Рабочее помещение, в котором присутствуют персональные компьютеры, должно удовлетворять эргономическим требованиям:

1. Помещение должно иметь естественное и искусственное освещение.

2. Рабочие места по отношению к световым проемам должны располагаться так, чтобы естественный свет падал сбоку, преимущественно слева.

3. Площадь на одно рабочее место пользователя персонального компьютера на базе электроннолучевой трубки должна составлять не менее 6 м², на базе плоских дискретных экранов (жидкокристаллические, плазменные) - 4,5 м².

4. При размещении рабочих мест с компьютерами расстояние между рабочими столами с мониторами должно быть не менее 2,0 м, расстояние между боковыми поверхностями видеомониторов – не менее 1,2 м.

5. Конструкция рабочего стола должна обеспечивать оптимальное размещение на рабочей поверхности используемого оборудования с учетом его количества и конструктивных особенностей, характера выполняемой работы.

6. Экран видеомонитора должен находиться от глаз пользователя на оптимальном расстоянии 600-700 мм, но не ближе 500 мм.

7. В помещениях с компьютерами ежедневно должна проводиться влажная уборка.

Оптимальное размещение предметов труда и документации в зонах досягаемости (рисунок 92.):

1. Дисплей размещается в зоне «а» (в центре).
2. Системный блок размещается в предусмотренной нише стола.
3. Клавиатура – в зоне «г/д».
4. «Мышь» – в зоне «в» справа.
5. Документация, необходимая при работе – в зоне легкой досягаемости ладони – «б», а в выдвижных ящиках стола – редко используемая литература.

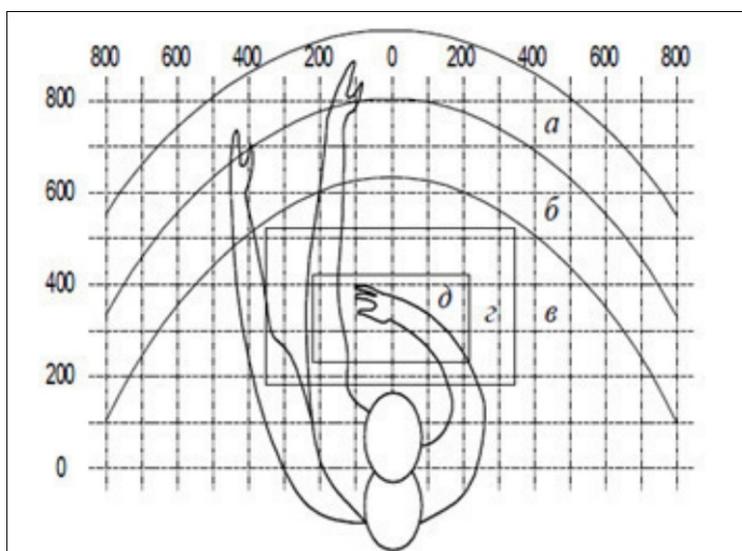


Рисунок 92 – Эргономические требования

(а – зона максимальной досягаемости; б – зона досягаемости пальцев при вытянутой руке; в – зона легкой досягаемости ладони; г – оптимальное пространство для грубой ручной работы; д – оптимальное пространство для тонкой ручной работы.).

6.4 Производственная безопасность

В соответствии с ГОСТ 12.0.002-2014 вредным фактором является фактор, воздействие которого на работающего в определенных условиях может привести к заболеванию, снижению работоспособности и (или)

отрицательному влиянию на здоровье потомства, а опасным производственным фактором – фактор, воздействие которого на работающего в определенных условиях приводит к травме, острому отравлению или другому внезапному резкому ухудшению здоровья, или смерти.

Наиболее существенными вредными факторами являются: отклонение показателей микроклимата рабочей зоны, недостаточная освещенность рабочей зоны, повышенный уровень шума, электромагнитные излучения, ионизирующее излучение. Также стоит учесть и опасные факторы, к данным факторам в первую очередь относятся: поражение электрическим током, а также опасные факторы, связанные с пожаром, например: огонь, дым, повышенная температура окружающей среды.

Возможные опасные и вредные факторы сведены в таблицу 18.

Таблица 18 – Возможные опасные и вредные факторы

Факторы (ГОСТ 12.0.003-2015)	Этапы работ		Нормативные документы
	Разработка	Эксплуатация	
Отклонение показателей микроклимата	+	+	ГОСТ 12.1.005-88 ССБТ. Общие санитарно-гигиенические требования к воздуху рабочей зоны. [66]
Недостаточная освещенность рабочей зоны	+	+	ГОСТ Р 55710-2013 Освещение рабочих мест внутри зданий. Нормы и методы измерений [74] ГОСТ Р 54944-2012. Здания и сооружения. Методы измерения освещенности [67].
Повышенный уровень шума	+	+	ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности [68].
Электромагнитные излучения	+	+	ГОСТ 12.1.006-84 ССБТ. Электромагнитные поля радиочастот. Общие требования безопасности [70].
Ионизирующее излучение	+	+	ГОСТ 15484-81 Излучения ионизирующие и их измерения, термины и определения [71].
Поражение электрическим током	+	+	ГОСТ 12.1.019-2017 ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты [72].
Опасные факторы, связанные с пожаром	+	+	ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования [73].

6.5 Анализ опасных и вредных факторов

6.5.1 Микроклимат помещения

Комфортность труда и высокая производительность работающего зависит от микроклимата в помещении. Микроклимат определяется действующими на организм человека сочетаниями температуры, относительной влажности и скорости движения воздуха. Оптимальные и допустимые значения характеристик микроклимата, согласно СанПиН 2.2.4.548–96, обеспечивающие комфортную работу трудящегося, не вызывают отклонений в состоянии здоровья, поддерживают высокий уровень работоспособности.

В таблице 19 приведены рекомендуемые параметры микроклимата.

Таблица 19 – Рекомендуемые параметры микроклимата

Период года	Категория работ по уровню энергозатрат, Вт	Температура воздуха, °С	Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	Ia (до 139)	(22-24)	(21-25)	(60-40)	0,1
Теплый	Ia (до 139)	(23-25)	(22-26)	(60-40)	0,1

Выполняемые работы относятся к категории Ia (работы с интенсивностью энергозатрат до 139 ккал/ч, производимые сидя и сопровождающиеся незначительным физическим напряжением). Следовательно, в рабочем помещении должна поддерживаться температура (22 – 24) °С в холодное время года, (23 – 25) °С в теплый период года.

Таким образом, для обеспечения заданного интервала температур необходимо обеспечить помещение системой кондиционирования, либо в летний период регулярно проветривать помещение, а в зимний использовать обогреватели.

6.5.2 Повышенный уровень шума

Важной характеристикой рабочего помещения является уровень шума. Шумом называют любой нежелательный звук или совокупность таких звуков. Длительное воздействие шума может отрицательно сказаться на здоровье работника, а в некоторых крайних случаях даже привести к глухоте.

Для описываемого помещения основными источниками шума можно считать:

1. шум систем охлаждения ПК;
2. уличный шум;
3. бытовые шумы.

Для выполняемых работ (легкая физическая нагрузка, напряженный труд 1 степени) уровень звука не должен превышать 60дБА. Для уменьшения шумов, генерируемых персональными компьютерами, рекомендуется проводить их регулярную диагностику, выполнять чистку и устранение неисправностей. Для устранения бытовых шумов может быть применена звукоизоляция.

6.5.3 Электромагнитные излучения

Источником электромагнитных излучений являются любые электроприборы, к которым относится и ЭВМ. Длительное воздействие электромагнитного поля на организм человека может вызвать нарушение функционального состояния нервной и сердечно-сосудистой систем, что выражается в повышенной утомляемости, снижении качества выполнения рабочих операций, сильных болях в области сердца, изменении кровяного давления и пульса.

Оценка опасности воздействия электромагнитного поля на человека производится по величине электромагнитной энергии, поглощаемой телом человека, с учетом электрической и магнитной напряженностей поля.

Практически при обслуживании даже мощных, электроустановок высокого напряжения магнитная напряженность значительно меньше опасной (в 8 раз), поэтому оценку потенциальной опасности воздействия электромагнитного поля достаточно производить по величине электрической напряженности поля. В соответствии с ГОСТ 12.1.002—84, ССБТ «Электромагнитные поля токов промышленной частоты. Общие требования безопасности», нормы допустимых уровней напряженности электромагнитных полей зависят от времени пребывания человека в контролируемой зоне. Присутствие персонала на рабочем месте в течение 8 ч допускается при напряженности, не превышающей 5 кВ/м. Основными видами средств коллективной защиты от воздействия электромагнитного поля токов промышленной частоты являются стационарные или переносные заземленные экранирующие устройства. Так как электромагнитное излучение в месте работы не превышает 5 кВ/м, применение экранирующих устройств не требуется.

6.5.4 Ионизирующее излучение

Основным источником ионизирующего излучения является дисплей компьютера (ЭВМ).

Ионизирующее излучение может вызывать торможение функций кроветворных органов, нарушение нормальной свертываемости крови и увеличение хрупкости кровеносных сосудов, снижение сопротивляемости организма к инфекционным заболеваниям.

Доза облучения при расстоянии до дисплея 20 см составляет 50 мкбэр/час. Конструкция ПЭВМ должна обеспечивать мощность экспозиционной дозы мягкого рентгеновского излучения в любой точке на расстоянии 0,05 м от экрана и корпуса не более 1 мкЗв/час (100 мкР/час).

Для защиты от внешнего облучения, возникающего при работе с дисплеем, проводятся следующие мероприятия:

1. для обеспечения оптимальной работоспособности и сохранении здоровья на протяжении рабочей смены должны устанавливаться регламентированные перерывы – при 8-часовом рабочем дне продолжительностью 15 минут через каждый час работы;

2. дисплей устанавливается таким образом, чтобы от экрана до сотрудника было не менее (60-70) см;

3. должны использоваться дисплеи со встроенными защитными экранами.

Для мониторов рекомендуется следующее дооснащение:

1. защитный фильтр для экрана, ослабляющий переменное электрическое и электростатическое поля;

2. для одиночных ЭВМ или их однорядном расположении – специальное защитное покрытие на переднюю панель и боковые стенки;

3. при многорядном расположении ЭВМ, если соседние рабочие места располагаются близко друг к другу (на расстоянии (1,2 - 2,5) м) – защитное покрытие задней и боковых стенок, монтирование специальных экранирующих панелей с задней и боковых сторон монитора, установка перегородок между различными пользователями.

6.5.5 Освещённость рабочей зоны

Освещение — неотъемлемый элемент условий трудовой деятельности человека.

При правильно организованном освещении рабочего места сохраняется зрение человека и нормальное состояние его нервной системы, а также обеспечивается безопасность в процессе производства.

Производительность труда и качество выпускаемой продукции находятся в прямой зависимости от освещения.

Рабочая зона освещается таким образом, чтобы можно было отчетливо видеть процесс работы, не напрягая при этом зрения.

Осветительные приборы и рабочее место располагаются таким образом, чтобы отсутствовало прямое попадание лучей источника света в глаза.

Уровень необходимого освещения определяется степенью точности зрительных работ. Наименьший размер объекта различения составляет 0,5 – 1 мм. В помещении присутствует естественное освещение. По нормам освещенности и отраслевым нормам, работа за ПК относится к зрительным работам средней точности для любого типа помещений. Нормирование освещенности для работы за ПК приведено в таблице 20.

Таблица 20 – Нормирование освещенности для работы с ПК

Характеристика зрительной работы	Наименьший или эквивалентный размер объекта различения, мм	Относительная продолжительность зрительной работы при направлении зрения на рабочую поверхность, %	Искусственное освещение		Естественное освещение	
			Освещенность на рабочей поверхности от системы общего освещения, лк	Коэффициент пульсации и освещенности КП, %, не более	КЕО, %, при	
					верхнем или комбинированном	боковых
Средней точности	От 0,5 до 0,1	Не менее 70	200	5	4	1,5
		Менее 70	150	10	4	1,5

Требования к освещению на рабочих местах, оборудованных ПК, представлены в таблице 21.

Таблица 21 – Требования к освещению на рабочих местах с ПК

Освещенность на рабочем столе	(200–400) лк
Освещенность на экране ПК	не выше 200 лк
Блики на экране	не выше 40 кд/м ²
Прямая блескость источника света	200 кд/м ²
Показатель ослепленности	не более 20
Показатель дискомфорта	не более 15
Отношение яркости между рабочими поверхностями	3:1–5:1

Отношение яркости между поверхностями стен и оборудования	10:1
Коэффициент пульсации:	не более 10%

6.5.6 Расчет освещенности рабочего места

Расчет освещения производится для комнаты площадью 20 м², ширина которого – 5 м, а длина – 4 м, рабочее место представлено на рисунке 93.

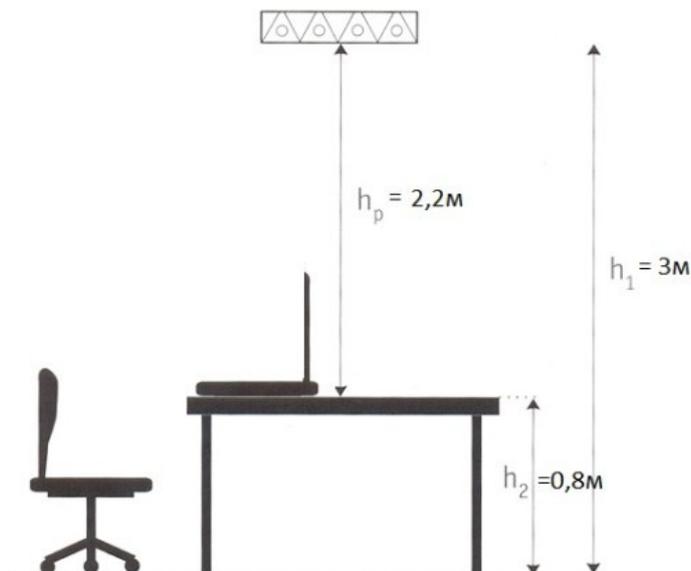


Рисунок 93 – Рабочее место

Определение количества светильников осуществляется методом светового потока.

Расчет производится под типовой растровый светильник с 4-ю линейными люминесцентными лампами (Кл), мощностью 18 Вт каждая (1 лампа даёт световой поток (СПл), равный 900 лк).

Для расчета коэффициента использования необходимо вычислить индекс помещения по формуле [44]:

$$I = \frac{S}{h \cdot (A + B)}, (44)$$

где S – площадь помещения, $S = 20 \text{ м}^2$;

h – расчетная высота подвеса, $h = 3 \text{ м}$;

A – ширина помещения, $A = 5 \text{ м}$;

B – длина помещения, $B = 4 \text{ м}$.

По имеющимся данным рассчитаем индекс помещения:

$$I = \frac{20}{(3-0,8) \cdot (5+4)} = 1,01. (45)$$

Согласно рисунку 94 значение коэффициента использования $U = 65$.

		Отражение поверхностей							
Индекс помещения	Потолок	80	80	80	70	50	50	30	0
	Стены	80	50	30	50	50	30	30	0
	пол	30	30	10	20	10	10	10	0
	0,6	53	38	32	37	35	31	31	27
	0,8	60	45	38	44	41	38	37	34
	1	65	51	43	49	46	43	42	38
	1,25	70	57	49	54	51	48	47	44
	2	76	66	56	61	57	55	54	51
	2,5	78	70	59	64	60	58	57	54
	3	80	73	62	67	62	60	59	57
	4	81	76	64	69	63	62	61	58
	5	82	78	65	70	65	64	62	60

Рисунок 94 – Индекс помещения

Расчет количества светильников ($K_{св.}$) для данного помещения [45]:

$$K_{св.} = \frac{100 \cdot E \cdot S \cdot K_z}{U \cdot K_l \cdot СП_l} = \frac{500 \cdot 20 \cdot 100 \cdot 1,2}{65 \cdot 4 \cdot 900} = 5 \text{ светильников}, (46)$$

где E – нормированная минимальная освещенность, Лк; S – площадь освещаемого помещения; K_z – коэффициент запаса, учитывающий уменьшение светового потока лампы в результате загрязнения светильников в процессе эксплуатации; U – коэффициент использования (определяется по таблице коэффициентов использования различных светильников).

Следовательно, для рабочего места достаточно установить 5 растровых светильников.

6.5.7 Электробезопасность

ПЭВМ и периферийные устройства являются потенциальными источниками опасности поражения человека электрическим током. При работе с компьютером возможен удар током при соприкосновении с токоведущими частями оборудования.

Рабочие места с ПЭВМ должны быть оборудованы защитным занулением. Подача электрического тока в помещение должна осуществляться от отдельного независимого источника питания. Необходима изоляция токопроводящих частей и ее непрерывный контроль. Должны быть предусмотрены защитное отключение, предупредительная сигнализация и блокировка.

Электрические изделия по способу защиты человека от поражения электрическим током подразделяются на пять классов: 0, 01, 1, 2, 3.

ЭВМ можно отнести к классу 01, то есть, к изделиям, имеющим рабочую изоляцию, элемент для заземления и провод без заземляющей жилы для присоединения к источнику питания. При начале работы с ЭВМ необходимо проверить герметичность корпуса, не открыты ли токоведущие части.

Помещение, в котором расположено рабочее место, относится к категории помещений без повышенной опасности, и соответствует установленным условиям:

1. напряжение питающей сети 220 В, 50 Гц;
2. относительная влажность воздуха 50%;
3. средняя температура около 24°C;
4. наличие непроводящего полового покрытия.

6.5.8 Пожарная безопасность

Помещения, в зависимости от характеристики используемых в производстве веществ и их количества, по пожарной и взрывной опасности подразделяются на категории А, Б, В, Г, Д в соответствии НПБ от 18.06.2003 г. №105-03 «Определение категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности».

Здание, где расположено рабочее место, относится к категории «Д» по степени пожарной опасности, так как в нем отсутствует обработка пожароопасных веществ, отсутствуют источники открытого огня. А стены здания и перекрытия выполнены из трудно сгораемых и несгораемых материалов (кирпич, железобетон, и др.).

При неправильной эксплуатации оборудования и коротком замыкании электрической сети может произойти возгорание, которое грозит уничтожением ПЭВМ, документов и другого имеющегося оборудования. Система вентиляции может стать источником распространения возгорания.

В качестве возможных причин пожара можно указать следующие:

1. короткие замыкания;
2. перегрузка сетей, которая ведет за собой сильный нагрев токоведущих частей и загорание изоляции.

Необходимо предусмотреть ряд профилактических мероприятий технического, эксплуатационного, организационного плана.

Организационные мероприятия предусматривают:

1. противопожарный инструктаж обслуживающего персонала;
2. обучение персонала правилам техники безопасности;
3. издание инструкций, плакатов, планов эвакуации.

Эксплуатационные мероприятия:

1. соблюдение эксплуатационных норм оборудования;
2. обеспечение свободного подхода к оборудованию.
3. содержание в исправности изоляции токоведущих проводников.

Технические мероприятия:

1. установка датчиков пожарной сигнализации, реагирующих на появление дыма;
2. наличие системы оповещения персонала в случае аварийных ситуаций;
3. соблюдение противопожарных мероприятий при устройстве электропроводок, оборудования, систем отопления, вентиляции и освещения.
4. Профилактический осмотр, ремонт и испытание оборудования.

6.6 Экологическая безопасность

6.6.1 Анализ воздействия на окружающую среду

Анализ воздействия на окружающую среду необходим для оценки возможности проведения разработки. В данной разработке могут быть выделены следующие потенциальные угрозы для окружающей среды:

1. загрязнение атмосферы;
2. загрязнение гидросферы;
3. загрязнение почв.

При выполнении данной разработки необходимо контролировать утилизацию отходов, к которым относятся печатные бумажные материалы, расходные части печатающих устройств, вышедшие из строя комплектующие ПК и периферийные устройства (в том числе их комплектующие).

Для защиты окружающей среды от перечисленных выше факторов необходимо соблюдать правила утилизации твердых отходов и комплектующих ПЭВМ:

1. сдавать бумажные отходы в специальные организации для дальнейшей их переработки;
2. обращаться в специальные организации для утилизации вышедших из строя комплектующих и расходных материалов.

6.7 Безопасность в чрезвычайных ситуациях

6.7.1 Перечень возможных чрезвычайных ситуаций на объекте

Существуют несколько возможных чрезвычайных ситуаций на объекте:

1. Пожар.
2. Взрыв (с последующим горением).
3. Внезапное обрушение сооружений.
4. Аварии на электроэнергетических системах.
5. Аварии на коммунальных системах жизнеобеспечения.

Основной чрезвычайной ситуацией, возможной при выполнении работы можно считать пожар.

Основными причинами возникновения пожара считаются:

1. использование неисправных электроприборов;
2. перегрузка сети;
3. курение не в специальных местах;
4. оставление электрических приборов без присмотра на долгое

время.

Пожар может нанести не только вред здоровью, но и материальный ущерб. Применимо к выполняемой работе в случае пожара могут быть уничтожены бумажные документы и\или электронные носители информации. Для защиты информации рекомендуется использовать облачные хранилища данных для данных и документов. Для исходных кодов программ рекомендуется использовать системы контроля версий.

При обнаружении пожара необходимо:

1. сообщить в пожарную охрану по телефону 01 или 112 (с мобильного телефона);
2. оповестить лиц, находящихся в здании, о пожаре;
3. предпринять действия, способные прекратить пожар;
4. при опасности поражения электрическим током отключить электроэнергию;

5. эвакуироваться.

Ликвидация последствий пожара в помещении любого назначения требует больших усилий. Проведение подобных работ требует определенной последовательности от специалистов компании, для того чтобы работы осуществлялись по намеченному графику. В первую очередь необходимо очистить помещение от мусора, который скопился после пожара. После этого можно приступать непосредственно к устранению следов пожара и удалению слоя копоти с поверхностей.

Для того чтобы ликвидация последствий пожара в помещении осуществлялась эффективнее и быстрее, специалисты компаний используют различное оборудование, такое как пеногенераторы и специальные мойки под давлением.

Выводы по разделу

В рамках раздела «социальная ответственность» были выявлены и проанализированы наиболее вероятные вредные и опасные производственные факторы, а также предложены мероприятия по снижению уровней их воздействия на работника. Была приведена аннотация с описанием рабочего места и перечнем выполненных работ. Произведен расчет освещенности рабочего места. Помимо этого, в подразделе «экологическая безопасность» рассматривается характер воздействия проектируемого решения на окружающую среду. Также рассмотрены наиболее возможные чрезвычайные ситуации на рабочем месте и алгоритм действий при их возникновении.

Заключение

В магистерской диссертации произведено исследование предиктивной модели двигателей технологического процесса производства БОПП-пленки. Собраны данные в виде временных рядов для разработки предиктивной модели, которые отображают основные параметры двигателя – скорость, ток, температура и момент. После чего подготовлены для проведения анализа основных методов (прогнозных и поиска аномалий) на которых базируется предиктивная модель. Совершен сравнительный анализ методов и выявлены лучшие для последующего их внедрения в технологический процесс. Как видно из результатов сравнительного анализа в качестве прогнозной части можно использовать практически любой, кроме простого и двойного экспоненциального сглаживания, а в качестве части для поиска аномалий лучше всего себя показывает DBSCAN.

В результате работы является внедрение в разработанный имитационный технологический процесс. Построены информационные потоки для передачи данных по протоколу OPC UA, организовано их хранение в базе данных временных рядов InfluxDB, а также визуализированы эти данные с помощью платформы онлайн-мониторинга Grafana.

В итоге, разработана предиктивная модель и архитектура для её внедрения, что позволяет использовать её в любом технологическом процессе, а в частности в технологическом процессе производства БОПП-пленки.

Список литературы

1. G. Beier, A. Ullrich, S. Niehoff, M. Reibig. How it is defined from a sociotechnical perspective and how much sustainability it includes – A literature review. // Journal of Cleaner Production 259 (2020). – 9.03.2020.
2. А. И. Власов, П. В. Григорьев, А. И. Кривошеин. Модель предиктивного обслуживания оборудования с применением беспроводных сенсорных сетей. // Надежность и качество сложных систем № 2 (22). – 2018. – С. 26-35.
3. Н. И. Шаханов, И. А. Варфоломеев, Е. В. Ершов, О. В. Юдина. Прогнозирование отказов оборудования в условиях малого количества поломок. // Вестник Череповецкого государственного университета. – 2016. – С.36-41.
4. В. В. Кузьмин, Д. С.Косов, А. Л. Новиков, А. В. Иващенко. Система прогнозирования отказов оборудования промышленных предприятий. // Надежность и качество сложных систем № 3 (11). – 2015 – С.87-90.
5. Р. А. Бобров. Нейросетевые модели для обнаружения аномалий в многомерных временных рядах. / Магистерская диссертация. – МГУ имени М. В. Ломоносова. – 2019.
6. А. А. Цыпин. Определение интенсивности отказов для систем, описываемых временными рядами. / Выпускная квалификационная работа. – МГУ имени М. В. Ломоносова. – 2019.
7. Н. А. Романов. Детектирование аномалий во временных рядах при помощи глубоких нейронных сетей. / Магистерская диссертация. – МГУ имени М. В. Ломоносова. – 2018
8. А. В. Чимров. Применение виртуальных агентов систем управления для обнаружения и ликвидации аварийных ситуаций на объектах НГО. / Магистерская диссертация. – НИ ТПУ. – 2019.

9. Н. И. Шкель. Разработка прогнозных моделей на базе искусственных нейронных сетей. / Бакалаврская работа. – НИ ТПУ. – 2019.
10. Е. Е. Белуха. Исследование применимости метода SSA для обработки временных рядов в нефтегазовой отрасли. / Бакалаврская работа. – НИ ТПУ. – 2019.
11. А. В. Лисаченко. Разработка программного обеспечения для прогнозирования сезонных временных рядов. / Бакалаврская работа. – НИ ТПУ. – 2019.
12. Д. А. Беспалов, А. С. Болдырев. Способ мониторинга с прогнозированием состояния датчиковых сетей на основе статистики // Известия ЮФУ. Технические науки. – 2017. – С.220-229.
13. А. С. Родионов, Э. В. Мухаметзянов, А. М. Курбангалиев, Р. Р. Исламгулов. Программно-аппаратный комплекс для мониторинга показаний датчиков // Electrical and data processing facilities and systems. № 1, v. 13. – 2017. – С.66-71.
14. SIMATIC Программируемый контроллер S7-1500 (Руководство). – URL: <https://www.siemens-pro.ru/components/s7-1500.htm>
15. Ultrasonic Sensors Measure Roll Diameter and Detect Web Break (Manual). – URL: https://files.pepperl-fuchs.com/webcat/navi/productInfo/doct/tdoct5773__eng.pdf?v=23-MAR-20
16. Преобразователи SINAMICS и двигатели SIMOTICS. – URL: <https://www.prosoft.ru/cms/f/465023/%D0%9F%D0%BE%D0%BB%D0%BD%D1%8B%D0%B9+%D0%BA%D0%B0%D1%82%D0%B0%D0%BB%D0%BE%D0%B3+SINAMICS+%D0%B8+SIMOTICS.pdf>
17. Powerblock SINAMICS G120. – URL: https://www.siemens-pro.ru/docs/sinamics_gv/g120/PM250_HIM_ru_ru-RU.pdf
18. SINAMICS G120 Преобразователь с управляющими модулями CU230P-2. – URL: <https://www.prosoft.ru/cms/f/465026/%D0%A0%D1%83%D0%BA%D0%BE%D0%B2%D0%BE%D0%B4%D1%81%D1%82%D0%B2%D0%BE+%D1%8D>

[%D0%BA%D1%81%D0%BF%D0%BB](#)

[%D1%83%D0%B0%D1%82%D0%B0%D1%86%D0%B8%D0%B8+SINAMICS+G120.pdf](#)

19. Ю. П. Лукашин. Адаптивные методы краткосрочного прогнозирования временных рядов / Ю. П. Лукашин. – 2017. – С. 416.

20. Документация библиотеки NumPy. – URL:

<https://numpy.org/devdocs/>

21. Документация библиотеки Pandas. – URL:

<https://pandas.pydata.org/docs/>

22. Документация библиотеки matplotlib. – URL: <https://matplotlib.org/>

[contents.html](#)

23. Документация библиотеки seaborn. – URL:

<https://seaborn.pydata.org/>

24. Документация библиотеки SciKit-learn. – URL: [https://scikit-](https://scikit-learn.org/stable/user_guide.html)

[learn.org/stable/user_guide.html](#)

25. Документация библиотеки datetime. – URL:

<https://docs.python.org/3/library/datetime.html>

26. Документация библиотеки statsmodels. – URL:

<https://www.statsmodels.org/stable/index.html>

27. Документация библиотеки Facebook prophet. – URL:

<https://facebook.github.io/prophet/>

28. Документация библиотеки Amazon Gluon TS. – URL:

<https://aws.amazon.com/ru/blogs/opensource/gluon-time-series-open-source-time-series-modeling-toolkit/>

29. Документация библиотеки Tsfresh. – URL:

<https://tsfresh.readthedocs.io/en/latest/>

30. Документация библиотеки Pmdarima. – URL: [https://alkaline-](https://alkaline-ml.com/pmdarima/index.html#)

[ml.com/pmdarima/index.html#](#)

31. Документация библиотеки Hmmlern. – URL:

<https://hmmlearn.readthedocs.io/en/latest/tutorial.html>

32. Документация библиотеки CatBoost. – URL: <https://catboost.ai/docs/concepts/about.html>
33. Документация библиотеки XGBoost. – URL: <https://xgboost.readthedocs.io/en/latest/>
34. Документация библиотеки Keras. – URL: <https://keras.io/guides/>
35. Ю. В. Сажин, А. В. Катунь, Ю. В. Сарайкин. Анализ временных рядов и прогнозирование / Учебник. – МГУ им. Н. П. Огарева. – 2013.
36. И. А. Чучуева. Модель прогнозирования временных рядов по выборке максимального подобия / Диссертация на соискание ученой степени кандидата технических наук. – МГТУ имени Н. Э. Баумана. – 2012.
37. R. G. Brown. Smoothing forecasting and prediction of discrete time series. – 1963. – С. 468.
38. P. R. Winters. Forecasting sales by exponentially weighted moving averages // Management Science. – 1960. – Vol. 6. – №3.
39. Pandit, Sudhakar M.; Wu, Shien-Ming. Time Series and System Analysis with Applications. – 2001. – С. 606.
40. Дж. Бокс, Г. Дженкинс. Анализ временных рядов, прогноз и управление. – 1973. – С. 195.
41. R. J. Hyndman, G. Athanasopoulos. Forecasting: principles and practice. – URL: <https://otexts.com/fpp2/> – 2013.
42. M. Stamp. A Revealing Introduction to Hidden Markov Models. – San Jose State University. – 2018.
43. Н. Дрейпер, Г. Смит. Прикладной регрессионный анализ. – М.: Издательский дом «Вильямс». – 2007. – С. 912.
44. Введение в разработку CatBoost. Доклад Яндекса. – URL: <https://habr.com/ru/company/yandex/blog/458790/>
45. T. Chen, C. Guestrin. XGBoost: A Scalable Tree Boosting System. – URL: <https://arxiv.org/pdf/1603.02754-2016>.

46. Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen¹, Weidong Ma¹, Qiwei Ye¹, Tie-Yan Liu¹. LightGBM: A Highly Efficient Gradient Boosting Decision Tree // 31st Conference on Neural Information Processing Systems (NIPS 2017). – 2017.
47. D. P. Kingma, J. L. Ba. Adam: A Method for Stochastic Optimization. URL: <https://arxiv.org/abs/1412.6980> – 2014.
48. Gábor Petneházi. Recurrent Neural Networks for Time Series Forecasting. URL: <https://arxiv.org/abs/1901.00069> – 2019.
49. J. Taylor, B. Letham. Forecasting at Scale // The American Statistician 72(1). – 2017.
50. D. Salinas, V. Flunkert, J. Gasthaus. DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. URL: <https://arxiv.org/abs/1704.04110> – 2019.
51. А. Г. Дьяконов, А. М. Головина. Выявление аномалий в работе механизмов методами машинного обучения // XIX Международная конференция «Аналитика и управление данными в областях с интенсивным использованием данных». – 2017.
52. M. Frigge, D. C. Hoaglin, B. Iglewicz. Some Implementations of the Boxplot // The American Statistician : journal. — 1989. — Vol. 43, no. 1. — P. 50—54.
53. М. Мельник. Основы прикладной статистики. — Москва: Энергоатомиздат, 1983. — 416 с.
54. ГОСТ Р ИСО 7870-2-2015 Статистические методы. Контрольные карты. Часть 2. Контрольные карты Шухарта. – 2016.
55. M. M. Breuning, H. Kriegel, T. Ng, J. Sander. LOF: Identifying Density-Based Local Outliers // Proc. ACM SIGMOD 2000 Int. Conf. On Management of Data, Dalles, TX. – 2000.
56. М. Ю. Чесноков. Поиск аномалий во временных рядах на основе ансамблей алгоритмов DBSCAN // Искусственный интеллект и принятие решений. – 2018. – С.98-106.

57. К. В. Воронцов. Математические методы обучения по прецедентам (теория обучения машин). – 2013.
58. Liu F. T., Ting K. M., Zhou Z. H. Isolation Forest // Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on. – IEEE, 2008. – С. 413-422.
59. Siemens TIA Portal. – URL: [siemens.com/global/en/products/automation/](https://www.siemens.com/global/en/products/automation/)
60. Документация Process Simulator. – URL: <https://automation.ucoz.com/>
61. Исходный код python-opcua. – URL: <https://github.com/FreeOpcUa/python-opcua>
62. Е. В. Шевчук, Ю. В. Шевчук. Современные тенденции в области хранения и обработки сенсорных данных. // Программные системы: теория и приложения №4(27). – 2015. – С.157-176.
63. Выбираем СУБД для хранения временных рядов. – URL: <https://habr.com/ru/company/oleg-bunin/blog/329062/>
64. Документация платформы Grafana. – URL: <https://grafana.com/docs/grafana/latest/>
65. Документация библиотеки influxdb. – URL: <https://influxdb-python.readthedocs.io/en/latest/index.html>
66. ГОСТ 12.1.005-88 ССБТ. Общие санитарно-гигиенические требования к воздуху рабочей зоны.
67. ГОСТ Р 54944-2012. Здания и сооружения. Методы измерения освещенности.
68. ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности.
69. Сапронов, Ю.Г. Аттестация рабочих мест по условиям освещённости: лабораторный практикум по дисциплине «Безопасность жизнедеятельности». – Шахты: ГОУ ВПО «ЮРГУЭС», 2010. – 64 с.

70. ГОСТ 12.1.006-84 ССБТ. Электромагнитные поля радиочастот. Общие требования безопасности.

71. ГОСТ 15484-81 Излучения ионизирующие и их измерения, термины и определения

72. ГОСТ 12.1.019-2017 ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты

73. ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования.

74. ГОСТ Р 55710-2013 Освещение рабочих мест внутри зданий. Нормы и методы измерений

75. Трудовой кодекс РФ.

Приложение А
(справочное)

Description of the manufacturing process of the BOPP film
TDO equipment
Scheme of information flows
Development of a predictive model algorithm

Студент:

Группа	ФИО	Подпись	Дата
8ТМ91	Поданев Максим Николаевич		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОАР ИШИТР	Заревич Антон Иванович	к.т.н		

Консультант – лингвист отделения иностранных языков ШБИП:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИЯ ШБИП	Сидоренко Татьяна Валерьевна	к.п.н		

1 Description of the manufacturing process of the BOPP film

BOPP film is a biaxially oriented polypropylene film. It is used for food products and everyday life products.

The production technology of the BOPP film (Appendix A) begins with the reception of granular raw materials at the discharge site. The granulate is loaded via pneumatic conveying into storage silos.

The film web consists of 5 layers. The central layer is formed by the main extruder, where the raw materials from the silos enter. The purpose of the main extruder is to ensure, due to high temperatures and mechanical impact, melting of the materials and homogenization of the melt in composition, temperature, pressure and viscosity. A pair of auxiliary satellite coextruders form intermediate layers and serve to give films additional properties mainly for appearance. The extreme layers are formed by two satellite extruders.

Next, the film enters the candle filter, where the melt entering the die is filtered. Inside the die, the melts supplied by the extruders go through independent collectors, where they merge into one common stream at a short distance in front of the exit slit. The total melt flow exits the head through a flat exit slit. The gap is regulated by 64 thermal bolts. Each thermal bolt is equipped with a heating element and a temperature sensor.

The preform film is formed on an injection machine consisting of an irrigation drum, a water bath, an air knife for pressing the film to the irrigation drum, a water cooling system for the irrigation drum and bath circuit, an air knife system and shafts for drying the film, an air knife for drying the irrigation drum. At the irrigation drum, the continuous polymer melt tape is cooled and crystallized. To obtain a homogeneous crystalline material, it is necessary that the melt is constantly cooled under the same conditions over the entire width of the drum. The air knife is used to dry the film.

After the irrigation drum complex, the first intermediate rewinder is installed in front of the longitudinal orientation machine, which is used when starting up and debugging equipment. In a machine of longitudinal orientation, the

process of drawing (orientation) of the film in the longitudinal direction is carried out. The machine includes a system of shafts with drives and instrumentation.

The longitudinally oriented film is fed through the guide rolls and the filling device to the transverse orientation unit of the TDO, where the edges of the film are captured by special clamps (clamps) 120 mm wide, located on two automatically lubricated closed chains (chain pitch 60 mm), moving on guide rails. The film moves through the pre-heating zone, where the film is heated with hot air. Then it moves through the stretch zone, where the film is stretched in the transverse direction. After that it moves through the heat-setting zone, where the internal stress in the film is removed and the desired appearance of the film is achieved.

From the exit of the TDO installation, the biaxially oriented film enters the system of the mill of pulling shafts (PRS), which includes all the nodes and devices from the output of the TDO to the rewinder. All shafts of the PRS system are divided into groups; they are rotated by DC drives and perform the following functions:

1. Group 1 selects the film from the TDO installation, aligns the web, cools and wraps the film through a thickness gauge. It includes leveling rollers, cooling shafts with internal water cooling.

2. Group 2 ensures the removal of cut off thickened edges (edges). It includes leveling rollers, pulling shaft.

3. Group 3 provides a flame treatment of the outer side of the film and the wiring of the film to the next group. It includes chilled shaft and pinch shaft.

4. Group 4 provides corona treatment of the outer side of the film and the wiring of the film to the next group. It includes a directing shaft the cooled dielectric shaft and a clamping shaft.

5. Group 5 provides corona treatment of the inner side of the film. It includes a cooled dielectric shaft and pinch shaft.

6. Clamping shafts prevent the film from slipping and removing air between the film and the shaft.

7. Group 6 provides cooling and film wiring to the rewinder. It includes internal water-cooled cooling shaft and guide shafts.

The film is given the properties required by the customer with the help of a gas-flame and corona treatment station.

After processing, the film is wound on sleeves. An overhead crane unloads them. Film samples are sent to the laboratory and quality control is in progress. Then, the finished film is sent to the warehouse.

In this master's work, TDO engines are studied, the incorrect operation of which leads to deformation of the film.

2 TDO equipment

2.1 Controlling equipment TDO

On the BOPP film production line, data collection and transmission is carried out by three leading PLCs. Each PLC controls its part of the production line. In the investigated installation of transverse tension (TDO), Siemens S7-300 is used (Fig. 5.) [14] with the following characteristics:

Table 1. Characteristics of the Siemens S7-300 controller

Operating temperature	Interfaces	Data transfer protocol	Discrete input / output Analog	Analog input / output	Operation time	MTBF	Price
-40°C ~ +70°C	Ethernet, RS232, RS485, USB	PROFINET IO, PROFIBUS DP, AS-Interface	10/6	8/4	0,1 mcs	350 000 h	400\$



Figure 5. Siemens S7-300

Siemens S7-300 is necessary for the construction of various kinds of automation systems (low and medium degree of complexity). The controller has a modular design and includes the following modules:

1. central processor module,
2. power supplies,
3. signal modules,
4. communication modules,
5. functional modules,
6. interface modules.

The controller design stands out for its high serviceability:

1. Each module is mounted on the S7-300 profile bus and fixed in working positions by screws. Using bus connectors, the modules are combined into one system.
2. The modules are placed randomly in the mounting racks. Due to the presence of removable front connectors.

The central processor Siemens S7-300 has the following indicators:

1. Performance (one of the highest controllers on the market).
2. Many active communications connections.
3. CPU 3xxC, CPU 31xT-2 DP have a set of built-in analog inputs / outputs. That is, the central processor can be used as a ready-made control unit.

Siemens S7-300 programming is implemented in accordance with IEC 61131-3 in the TIA Portal environment.

2.2 Sensors

To control film breakage, ultrasonic breakage sensors UC-F77 (Fig. 6.) are used [15].



Figure 6. A break sensor

Sensors allow accurate detection of complex materials such as transparent film. They are able to adjust the width of the sound beam and ignore unwanted reflections. They have high noise immunity.

2.3 Motors

To pass the film through the TDO, two asynchronous motors are used at the input of the unit and two motors at the output – SIEMENS SIMOTICS M-1PH8 (Fig. 7.) [16].



Figure 7. M-1PH8

Motors are synchronized using Siemens encoders and frequency converters SINAMICS S120.

2.4 Rectifier

Siemens Powerblock converts alternating current into direct current for motor control (Fig. 8.) [17].



Figure 8. A rectifier

2.5 Control Unit

The engines are controlled by the SINAMICS S120 drive system (Fig. 9.) [18]. The control module calculates the drives and the mains supply / recovery. Power units (power modules, power modules and engine modules) provide optimal energy conversion for both engines.



Figure 9. SINAMICS S120

“TDO” and “Slave” SINAMICS S120 are installed on the TDO. “Slave” is used to control the speed and torque of the engines. In addition, they process information about the settings, turning on and off the engine. "Master" synchronizes the operation of the "Slave" devices.

3 Scheme of information flows

The flow chart is presented in Appendix B. This flow chart consists of four levels:

1. The level of sensors and actuators. Level Interfaces – 1-WIRE, 4-20Ma, DRIVE-CLIQ. Devices - sensors (fixing film rupture) and actuators (engines). Signals - analog and discrete.

2. The controller level. Interface – RS485. Protocols – PROFIBUS, PROFINET. Devices – PLC (Siemens S7-300) and input / output modules. Three PLCs collect information along the entire technological line with the help of input / output modules from sensors; depending on it, control signals are generated for actuators. In addition, they transmit data to the control level.

3. Dispatch level. Interface – RJ-45. Protocols – Profinet, Ethernet, Modbus TCP / IP. Devices – personal computers, network switches, local HMI panels. Information from the PLC enters the SCADA package (Vijeo Citect) installed on the operator's workstation. SCADA-package allows you to visualize the operation of the production line, archive its data, as well as manage it.

4. SMSS level. At this level, time series are archived with a database, forecasts are generated, and anomalies are detected.

4 Development of a predictive model algorithm

4.1 Steps in developing a predictive model

1. Statement of the problem. Determining how the predictive model will be used, what they are needed for, its components and how effective the integration of the model into production is.

2. Collection of information. The more historical data is accumulated, the more models for forecasting can be applied and the more effective they can be.

3. Preliminary analysis. Graphing the process under study to determine any patterns, trends, or outliers. Calculation of the correlation between the data to determine how strong the relationship is between them.

4. Building a predictive model. For predictive maintenance, it is necessary to carry out forecasts for hours / days / weeks in advance.

5. Building a model for detecting anomalies. After receiving the forecast, it is necessary to determine how stable the future process is.

6. The choice of models. Comparison of models depending on quality metrics, on the speed of work and the possibility of integration into production.

7. Integration into the technological process. Determining the flow of information for organizing data storage, training models on this data and presenting information in a form convenient for dispatchers / operators.

4.2 Data collection for the predictive model

To implement the predictive model, it is necessary to have a data set of the studied objects of the technological process, analyze it and begin development. Since the object of study is a transverse tension device, the information is collected by the TDO PLC and sent to the Vijeo Citect SCADA package.

Vijeo Citect SCADA package – Schneider Electric software for monitoring and control. Backs up all system components. It has a distributed client-server architecture. It stores the three-month data of the production line, which will be needed for research.

Data is stored in .hst format. But analysis requires a .csv table format. Therefore, using the Citect Trend Reader v1.2.8 program, the format is converted.

4.3 Statement of the problem

A time series is a set of observations obtained sequentially in time, usually at regular intervals [19].

Information with SCADA is represented by a multidimensional (or one-dimensional) time series $X = \{x_1, x_2, \dots, x_T\}$, $x_t \in R^m$, where each element is an m -dimensional vector representing the readings of m sensors and actuators at a time t . It is necessary to make a forecast from the point $\hat{x}_t(t)$ of the quantity $x_{t+\tau}$. At each point in time of the forecast, it determines whether the values from the sensors are abnormal or not.

The objects under investigation are engines at the inlet of the TDO and at the outlet. Speed, temperature, load data for each of the objects.

4.4 Time series requirements

To make operational forecasts using autoregressive models, the time series should be stationary. According to the stationary series, it is possible to build a forecast since its future statistical characteristics will not differ from the observed current ones. The stationary process ξ_t is understood as such for which the following conditions are satisfied:

$$M[\xi_t] = \text{const} = a, \quad (1)$$

$$D[\xi_t] = \sigma^2, \quad (2)$$

$$r(t_i, t_j) = r(\tau), \quad (3)$$

where $M[\xi_t]$ is the mathematical expectation, $D[\xi_t]$ is the variance, $r(t_i, t_j)$ is the autocorrelation function, $\tau = t_i - t_j$.

The magnitude of the autocorrelation function depends on the magnitude of the shift τ , the value of a determines a constant level relative to which the values of the time series vary, and the constant value σ^2 characterizes the magnitude of this variation.

To test for stationarity, the Dickey-Fuller unit root test is used. An assumption is made about the type of process that generated this time series. An auxiliary model is constructed and hypotheses about the coefficients of this model are tested. Then the conclusion about the stationarity / non-stationarity of the original series is made.

Two hypotheses are put forward: the null hypothesis (H0) – there is a unit root, that is, the series is unsteady; alternative hypothesis (H1) – the null

hypothesis is rejected, the time series does not have a unit root, which means it is stationary. The output is interpreted using p-value. If the p-value is greater than 0.05, then the null hypothesis is not rejected and the series is non-stationary. If the p-value is less than or equal to 0.05, then the null hypothesis is rejected, which implies that the series is stationary.

In Figures 10, 11, 12, 13, time series of motor speed, current, temperature, and torque are checked for stationarity, autocorrelation, and partial autocorrelation.

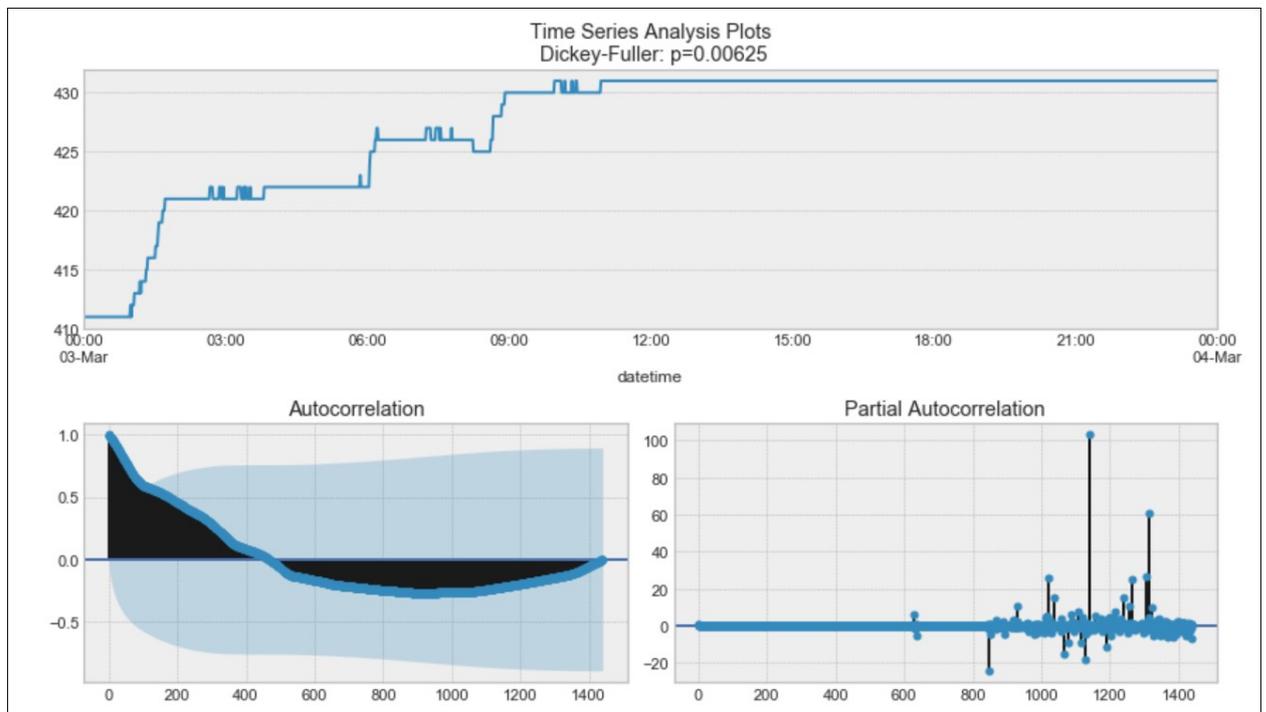


Figure 10. Time series analysis plots (SD)

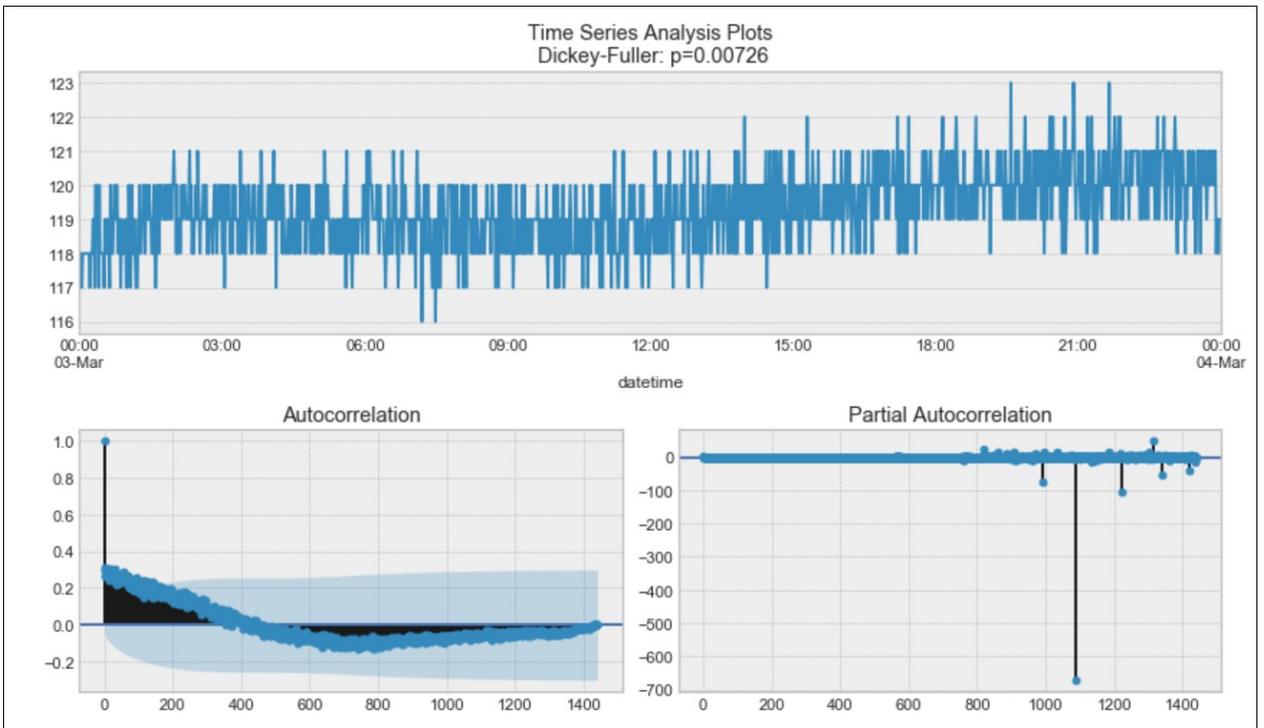


Figure 11. Time series analysis plots (CU)

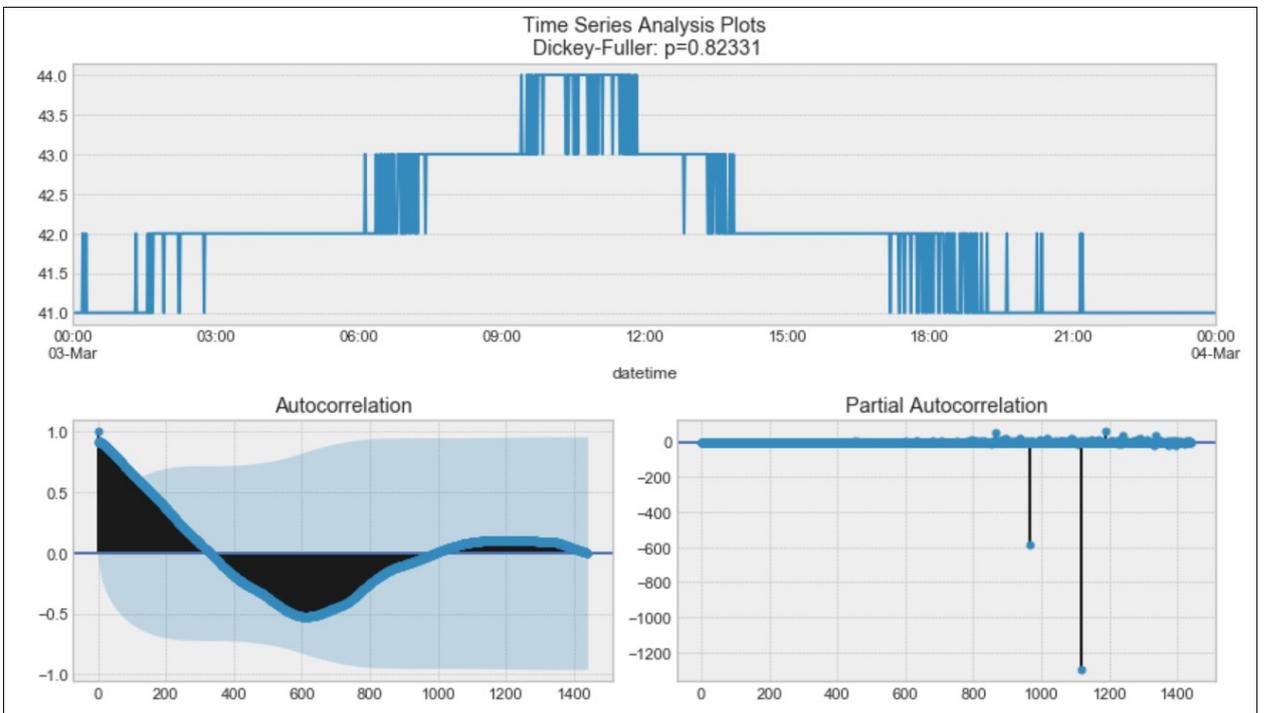


Figure 12. Time series analysis plots (TE)

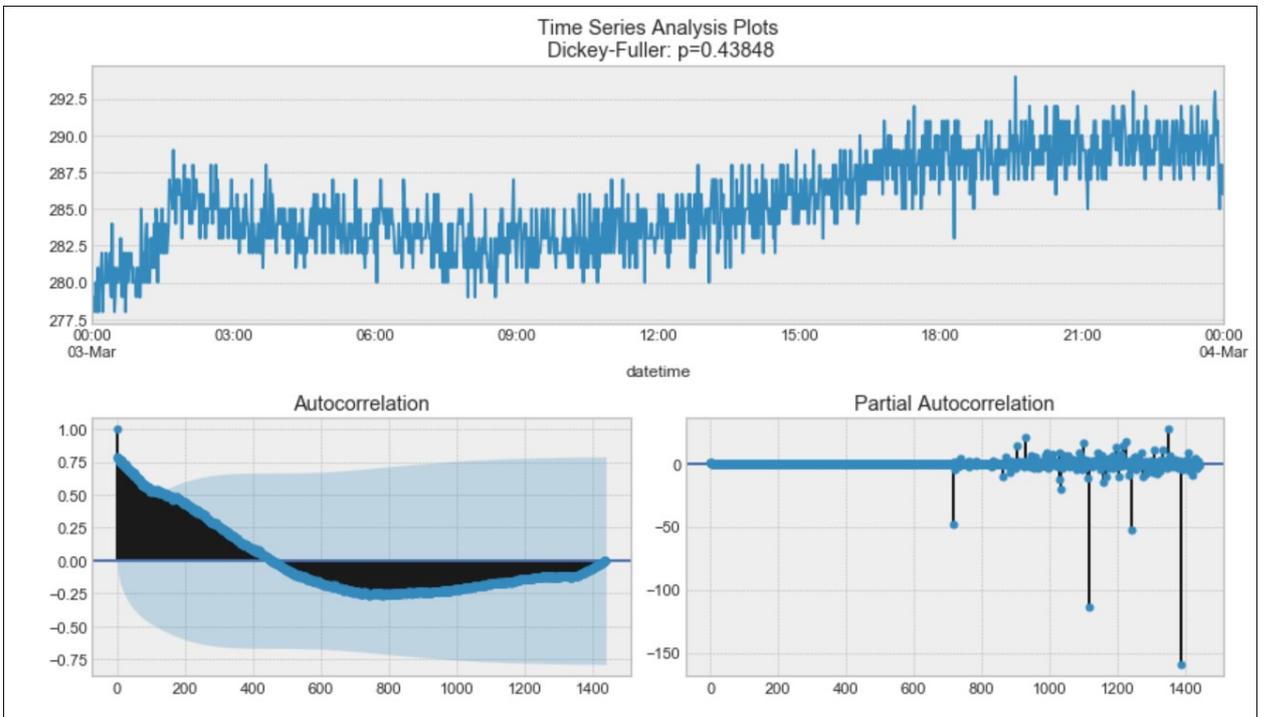


Figure 13. Time series analysis plots (NM)

Time series of temperature and engine torque do not pass the stationary test. Accordingly, it is necessary to carry out a differentiation operation, which involves subtracting a series from itself with a lag in one step:

$$y_1, \dots, y_T \rightarrow \dot{y}_2, \dots, \dot{y}_T, \quad (4)$$

$$\dot{y}_2 = y_t - y_{t-1}, \quad (5)$$

Stationarity tests after conversion are shown in Figures 14 and 15:

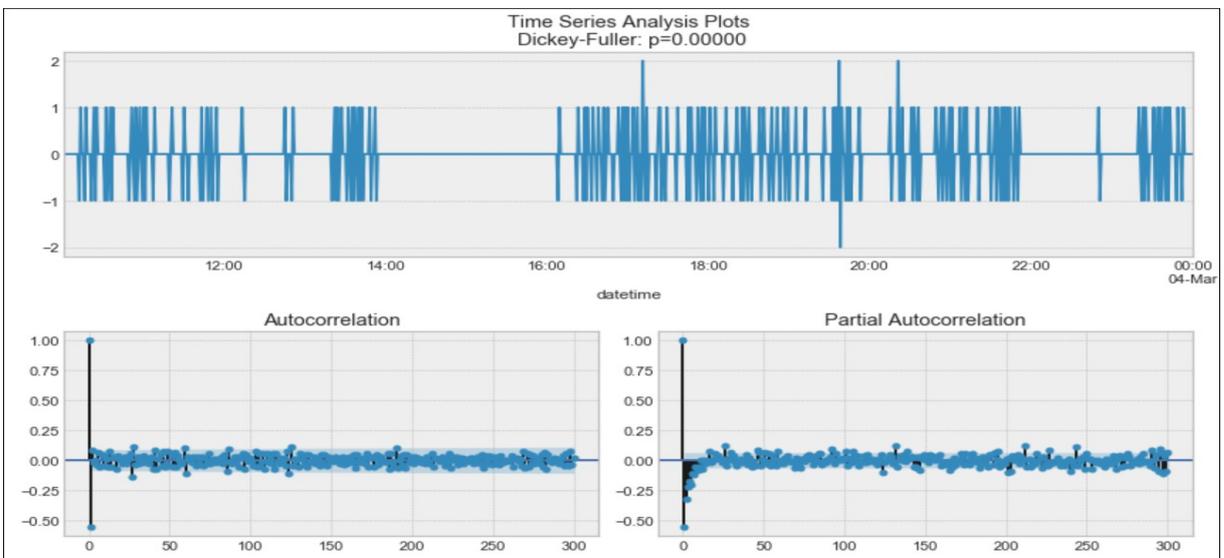


Figure 14. Time series analysis plots (TE)

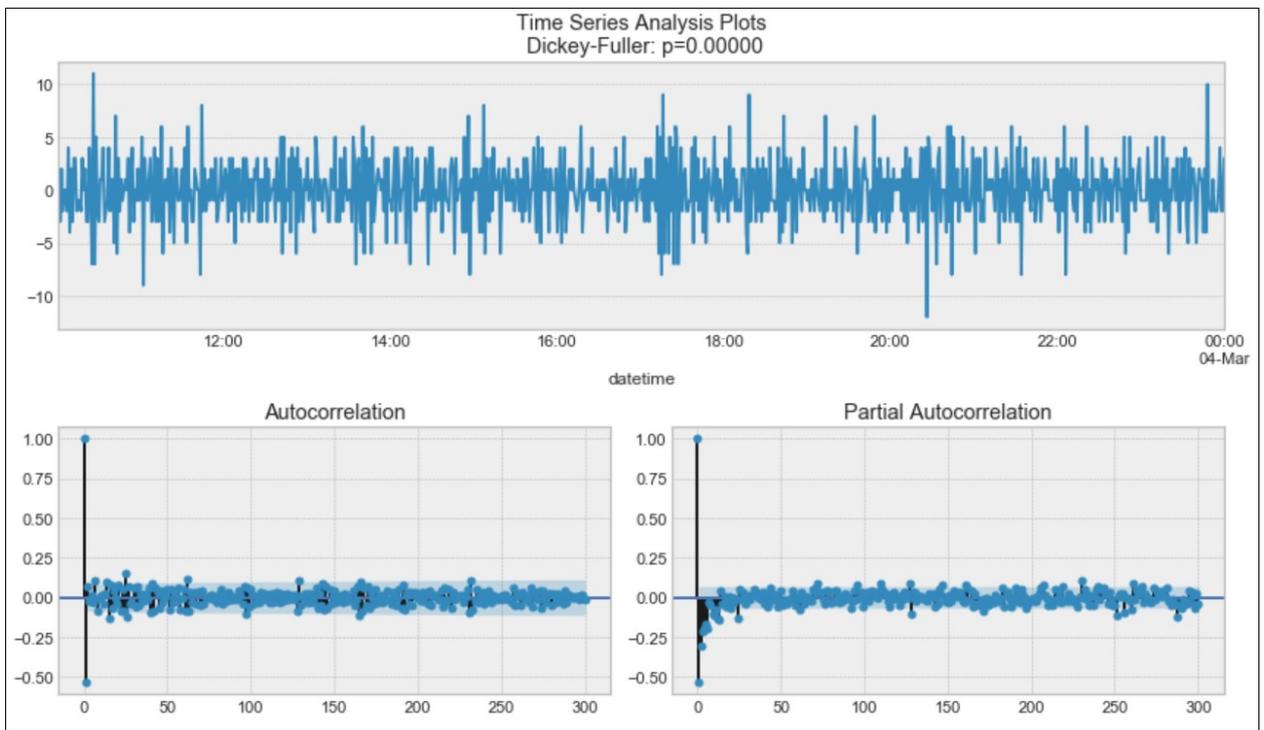


Figure 15. Time series analysis plots NM

4.5 Technology research stack

For the analysis of time series, there are several classic high-level programming languages – Matlab, R, Python, Julia. In cases where the solution of the problem involves the statistical processing of data, work with charts, etc., these languages are used more often than others.

This study uses Python, the web shell for IPython, the Jupyter Notebook, which combines code, text, and images. The analysis is carried out by the following libraries:

1. NumPy – a module for complex calculations. Provides functionality comparable to Matlab basic methods for manipulating large arrays and matrices [20].

2. Pandas - a library for processing and analyzing data. Built on top of the NumPy library. Provides special data structures and operations for manipulating numerical tables and time series – Series, DataFrame [21].

3. Matplotlib – a library for data visualization, both two-dimensional graphics and three-dimensional [22].

4. Seaborn – a data visualization library based on matplotlib. It allows using more types of graphs regarding matplotlib [23].
5. SciKit-learn – the library provides an implementation of a number of machine learning algorithms [24].
6. Datetime – the library is used to work with dates and times [25]
7. Statsmodels – a library that provides classes and functions for evaluating various statistical models, as well as for conducting statistical tests and researching statistical data [26].
8. Facebook prophet – a library designed to predict a large number of different business indicators. It does not require analysts to have deep knowledge of predictive model design [27].
9. Amazon GluonTS is a library developed by Amazon for time series analysis. There are tools necessary for building and training the most common neural network architectures, mechanisms for loading and data preprocessing, ready-made forecast models, and tools for evaluating and comparing various models [28].
10. Tsfresh – a library for generating features. For example, maximum, minimum values, p-value, slope coefficients of a trend line, etc. [29].
11. Pmdarima – a library that includes the equivalent of the auto.arima function from R, a collection of statistical tests of stationarity and seasonality, various operations – differentiation and inverse differentiation of time series, built-in sets of time series, and much more [30].
12. Hmmllearn – a library for using hidden Markov models [31].
13. CatBoost – a library of implementation of gradient boosting on open-source decision trees from Yandex. The algorithm used for ranking and forecasting, as well as recommender technologies [32].
14. XGBoost – a library for forecasting, classification, which is based on the gradient boosting algorithm of decision trees [33].
15. Keras – machine learning library. It is used to solve computer vision problems, natural language processing, time series forecasting, etc. [34].

4.6 Quality metrics

The main quality characteristics of the statistical models selected for forecasting are accuracy indicators. The concept of accuracy characterizes the degree of proximity of simulated (calculated by a certain forecast model) values of \hat{y} in their totality to the initial actual data y . The smaller the differences between theoretical values and empirical data, the higher the quality of the model [35].

Table 2. the most used quality metrics

Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \sum_{i=1}^n y_i)^2}$
Mean Absolute Deviation	$MAE = \frac{\sum_{i=1}^n y_i - \hat{y}_i }{n}$
Mean Standard error	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
Mean Absolute Percentage Error	$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{ y_i - \hat{y}_i }{y_i}$

Conclusion

The historical data of the BOPP film production process presented in the form of time series satisfy the necessary requirements. Time series are stationary and they do not have autocorrelation. The next step is to build forecasts and search for anomalies in the forecasts as well as to build forecasts, use the Python language, the necessary libraries and quality metrics.

Models for predicting and searching for anomalies are integrated into production on a production line and perform the functions of expert systems to increase the efficiency of production operators.

Приложение Б

Приложение В

```
from sklearn.metrics import mean_squared_error
from math import sqrt
import numpy as np
import pystan
import pandas as pd
from catboost import CatBoostRegressor

import matplotlib.pyplot as plt
import datetime
from sklearn.model_selection import TimeSeriesSplit
from dateutil.relativedelta import relativedelta
from scipy.optimize import minimize

import statsmodels.formula.api as smf
import statsmodels.tsa.api as smt
import statsmodels.api as sm
import scipy.stats as scs

from itertools import product
from tqdm import tqdm_notebook

import warnings
warnings.filterwarnings('ignore')

%matplotlib inline

from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import cross_val_score, TimeSeriesSplit
tscv = TimeSeriesSplit(n_splits=5)
scaler = StandardScaler()

def mean_absolute_percentage_error(y_true, y_pred):
    return np.mean(np.abs((y_true - y_pred) / y_true))*100

def format_data(X, column_name, start_drop=86401, end_drop=1048503):
    data = pd.read_csv(X, names=['4', '1', '2', '3'])
    data = data.drop(['3', '2', '4'], axis=1)
    data.rename(columns={'1':column_name}, inplace=True)
    data.drop(data.index[start_drop:end_drop], inplace=True)
    modern_data = pd.to_datetime(data.index)
    data_ = pd.DataFrame(modern_data, columns=['date'])
    data_['datetime'] = pd.to_datetime(data_['date'])
    data_ = data_.set_index('datetime')
    data_['date'] = data[column_name]
    data = data_.rename(columns={'date':column_name})
    data_ts = data.resample('T').ffill()
    return data_ts

data_ts_sd_m = format_data('datasets/Motor speed.csv', 'Motor speed')
data_ts_cu_m = format_data('datasets/Current.csv', 'Current')
data_ts_te_m = format_data('datasets/Temperature.csv', 'Temperature')
data_ts_nm_m = format_data('datasets/Torque.csv', 'Torque')

data_ts_sd_m['Current'] = data_ts_cu_m['Current']
data_ts_sd_m['Temperature'] = data_ts_te_m['Temperature']
data_ts_sd_m['Torque'] = data_ts_nm_m['Torque']
```

Продолжение приложения В

```
def timeseries_train_test_split(X, y, test_size):

    test_index = int(len(X)*(1-test_size))

    X_train = X.iloc[:test_index-1]
    y_train = y.iloc[:test_index-1]
    X_test = X.iloc[test_index:]
    y_test = y.iloc[test_index:]

    return X_train, X_test, y_train, y_test

def catboost_regression(column):
    y = data_ts_sd_m.dropna()[column]
    X = data_ts_sd_m.dropna().drop([column], axis=1)

    X_train, X_test, y_train, y_test = timeseries_train_test_split(X, y, test_size=0.3)

    X_train_scaled = scaler.fit_transform(X_train)
    X_test_scaled = scaler.transform(X_test)

    catboost = CatBoostRegressor(iterations=200, depth=16, learning_rate=0.5,
                                  loss_function='RMSE', verbose=0)

    catboost.fit(X_train_scaled, y_train, verbose=0)

    prediction = catboost.predict(X_test_scaled)
    error = mean_absolute_percentage_error(prediction, y_test)
    forecast = pd.Series(prediction, index=y_test.index)

    return forecast, error

forecast, error = catboost_regression('Motor speed')
forecast_2, error_2 = catboost_regression('Current')
forecast_3, error_3 = catboost_regression('Temperature')
forecast_4, error_4 = catboost_regression('Torque')
```

Приложение Г

```
def plotModelResults(model, X_train=X_train, X_test=X_test, plot_intervals=False, plot_anomalies=False):
    """
    Plots modelled vs fact values, prediction intervals and anomalies

    """

    prediction = model.predict(X_test)
    plt.figure(figsize=(20, 10))
    plt.tick_params(axis='both', which='major', labelsize=20)

    plt.title("Изменение скорости двигателя", fontsize=20)
    plt.xlabel('Время (м)', fontsize=20)
    plt.ylabel('Скорость двигателя (М/MN)', fontsize=20)

    plt.plot(y_test values, label="actual", linewidth=2.0)
    plt.plot(prediction, "g", label="prediction", linewidth=2.0)

    if plot_intervals:
        cv = cross_val_score(model, X_train, y_train,
                             cv=5,
                             scoring="neg_mean_absolute_error")
        mae = cv.mean() * (-1)
        deviation = cv.std()

        scale = 1.96
        lower = prediction - (mae + scale * deviation)
        upper = prediction + (mae + scale * deviation)

        plt.plot(lower, "r--", label="upper bond / lower bond", alpha=0.5)
        plt.plot(upper, "r--", alpha=0.5)

    if plot_anomalies:
        anomalies = np.array([np.NaN]*len(y_test))
        anomalies[y_test<lower] = y_test[y_test<lower]
        anomalies[y_test>upper] = y_test[y_test>upper]
        plt.plot(anomalies, "o", markersize=10, label = "Anomalies")

    error = rmse(y_test, prediction)
    plt.title("Поиск аномалий моделью LR", fontsize=20)
    plt.legend(loc="best")
    plt.tight_layout()
    plt.grid(True)
```

Приложение Д

```
from opcu import Client
import time
client = Client("opc.tcp://192.168.0.1:4840/")
client.connect()

while True:

    Motor_speed = client.get_node('ns=3;s="Motor speed"')
    Motor_speed = Motor_speed.get_value()
    print(Motor_speed)

    Current = client.get_node('ns=3;s="Current"')
    Current = Current.get_value()
    print(Current)

    Temperature = client.get_node('ns=3;s="Temperature"')
    Temperature = Temperature.get_value()
    print(Temperature)

    Torque = client.get_node('ns=3;s="Torque"')
    Torque = Torque.get_value()
    print(Torque)

    time.sleep(1)
```

Приложение Е

