

Школа: Инженерная школа информационных технологий и робототехники  
 Направление подготовки: Информационные системы и технологии  
 Отделение школы (НОЦ): Отделение информационных технологий

**БАКАЛАВРСКАЯ РАБОТА**

Тема работы
<b>КРАТКОСРОЧНОЕ ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ ГИБРИДНЫХ МЕТОДОВ</b>

УДК 004.421.5:519.216:519.246

Студент

Группа	ФИО	Подпись	Дата
8И6А	Репина Елизавета Юрьевна		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Кочегурова Е.А.	к.т.н.		

**КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:**

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН	Подопригора И.В.	к. э. н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ООД	Немцова О.А.			

**ДОПУСТИТЬ К ЗАЩИТЕ:**

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата

## РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Код результатов	Результат обучения (выпускник должен быть готов)
<b><i>Профессиональные и общепрофессиональные компетенции</i></b>	
P1	Применять базовые и специальные естественнонаучные и математические знания для комплексной инженерной деятельности по созданию, внедрению и эксплуатации геоинформационных систем и технологий, а также информационных систем и технологий в бизнесе.
P2	Применять базовые и специальные знания в области современных информационных технологий для решения инженерных задач.
P3	Ставить и решать задачи комплексного анализа, связанные с созданием геоинформационных систем и технологий, информационных систем в бизнесе, с использованием базовых и специальных знаний, современных аналитических методов и моделей.
P4	Выполнять комплексные инженерные проекты по созданию информационных систем и технологий, а также средств их реализации (информационных, методических, математических, алгоритмических, технических и программных).
P5	Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретация полученных данных, в области создания геоинформационных систем и технологий, а также информационных систем и технологий в бизнесе.
P6	Внедрять, эксплуатировать и обслуживать современные геоинформационные системы и технологии, информационные системы и технологии в бизнесе, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья, безопасность труда, выполнять требования по защите окружающей среды.
<b><i>Универсальные (общекультурные) компетенции</i></b>	
P7	Использовать базовые и специальные знания в области проектного менеджмента для ведения комплексной инженерной деятельности.
P8	Осуществлять коммуникации в профессиональной среде и в обществе в целом. Владеть иностранным языком (углублённый английский язык), позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты комплексной инженерной деятельности.
P9	Эффективно работать индивидуально и в качестве члена команды, состоящей из специалистов различных направлений и квалификаций.
P10	Демонстрировать личную ответственность за результаты работы и готовность следовать профессиональной этике и нормам ведения комплексной инженерной деятельности
P11	Демонстрировать знания правовых, социальных, экологических и культурных аспектов комплексной инженерной деятельности, а также готовность к достижению должного уровня физической подготовленности для обеспечения полноценной социальной и профессиональной деятельности.

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: Информационные системы и технологии  
 Отделение школы (НОЦ): Отделение информационных технологий

УТВЕРЖДАЮ:  
 Руководитель ООП  
 \_\_\_\_\_ Кочегурова Е.А.  
 (Подпись)    (Дата)    (Ф.И.О.)

**ЗАДАНИЕ**  
**на выполнение выпускной квалификационной работы**

В форме:

<b>Бакалаврской работы</b> <small>(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)</small>
---

Студенту:

Группа	ФИО
8И6А	Репиной Елизавете Юрьевне

Тема работы:

Краткосрочное прогнозирование временных рядов на основе гибридных методов	
Утверждена приказом директора (дата, номер)	№59-66/с от 22.02.2020

Срок сдачи студентом выполненной работы:	06.06.2020
--	------------

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ:**

<b>Исходные данные к работе</b>	<ul style="list-style-type: none"> <li>- гибридный метод прогнозирования на основе штрафного P-сплайна;</li> <li>- эволюционные методы оптимизации параметров модели;</li> <li>- методы кластеризации временных рядов.</li> </ul>
<b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b>	Обзор существующих моделей краткосрочного прогнозирования временных рядов, обзор методов оптимизации основных параметров модели. Оценка качества прогнозирования посредством вычисления ошибки прогноза и ее сравнения с результатами существующих методов прогнозирования.

<b>Перечень графического материала</b>	Презентация в формате *.pptx
<b>Консультанты по разделам выпускной квалификационной работы</b>	
<b>Раздел</b>	<b>Консультант</b>
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Подопригора И.В.
Социальная ответственность	Немцова О.А.
<b>Названия разделов, которые должны быть написаны на русском и иностранном языках:</b>	
Заключение	

<b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b>	01.03.2020
---	------------

**Задание выдал руководитель:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Кочегурова Е.А.	к.т.н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8И6А	Репина Елизавета Юрьевна		

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: Информационные системы и технологии  
 Отделение школы (НОЦ): Отделение информационных технологий  
 Период выполнения: весенний семестр 2019 /2020 учебного года

Форма представления работы:

<b>Бакалаврская работа</b>
----------------------------

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

### КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	06.06.2020
--	------------

Дата Контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
06.06.2020	Основная часть	75
18.04.2020	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	15
07.04.2020	Социальная ответственность	10

**СОСТАВИЛ:**

**Руководитель ВКР**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Кочегурова Е.А.	к.т.н.		

**СОГЛАСОВАНО:**

**Руководитель ООП**

Должность	ФИО	Ученая степень, звание	Подпись	Дата

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

Студентке:

Группа	ФИО
8И6А	Репина Елизавета Юрьевна

Школа	ИШИТР	Отделение школы (НОЦ)	Информационных технологий
Уровень образования	Бакалавриат	Направление/специальность	09.03.02 Информационные системы и технологии

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

1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	Работа с информацией, представленной в российских и иностранных научных публикациях, аналитических материалах, статистических бюллетенях и изданиях, нормативно-правовых документах.
2. <i>Нормы и нормативы расходования ресурсов</i>	
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	

**Перечень вопросов, подлежащих исследованию, проектированию и разработке:**

1. <i>Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения</i>	Оценка потенциальных потребителей, анализ конкурентных решений, QuaD и SWOT-анализ научного исследования, определение возможных альтернатив с использованием морфологического подхода.
2. <i>Планирование и формирование бюджета научных исследований</i>	Планирование этапов работ, определение трудоемкости работы, построение календарного графика (диаграмма Ганта), расчет бюджета научно-технического проекта.
3. <i>Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования</i>	Оценка ресурсной, финансовой и экономической эффективности научноисследовательского проекта.

**Перечень графического материала (с точным указанием обязательных чертежей):**

1. <i>Оценка конкурентоспособности технических решений</i>
2. <i>Матрица SWOT</i>
3. <i>Альтернативы проведения НИ</i>
4. <i>График проведения и бюджет НИ</i>
5. <i>Оценка ресурсной, финансовой и экономической эффективности НИ</i>

<b>Дата выдачи задания для раздела по линейному графику</b>	28.02.2020
---	------------

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН	Подопригора И.В.	к. э. н.		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8И6А	Репина Е.Ю.		

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студентке:

<b>Группа</b>	<b>ФИО</b>
8И6А	Репина Елизавета Юрьевна

<b>Школа</b>	<b>ИШИТР</b>	<b>Отделение (НОЦ)</b>	<b>информационных технологий</b>
<b>Уровень образования</b>	Бакалавриат	<b>Направление/специальность</b>	09.03.02 Информационные системы и технологии

Тема ВКР:

Краткосрочное прогнозирование временных рядов на основе гибридных методов

**Исходные данные к разделу «Социальная ответственность»:**

1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Разработка программного приложения для краткосрочного прогнозирования временных рядов на основе штрафного Р-сплайна. Приложение предназначено для применения на компьютерах с ОС Windows.
--	---

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

<b>1. Правовые и организационные вопросы обеспечения безопасности:</b> 1.1 специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; 1.2 организационные мероприятия при компоновке рабочей зоны.	Требования к организации и оборудованию рабочих мест с ПЭВМ, правовые нормы трудового законодательства
<b>2. Производственная безопасность:</b> 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	Вредные: 1. Отклонение показателей микроклимата в помещении; 2. Недостаточная освещенность рабочей зоны. Опасные: 1. Опасность поражения электрическим током.
<b>3. Экологическая безопасность:</b>	Выявление и анализ негативно влияющих на окружающую среду факторов при эксплуатации персонального компьютера. Определение решения по обеспечению экологической безопасности согласно нормативным документам.
<b>4. Безопасность в чрезвычайных ситуациях:</b>	Перечень возможных чрезвычайных ситуаций, которые могут возникнуть при работе в офисном помещении. Способы защиты от пожара и ликвидация последствий.

Дата выдачи задания для раздела по линейному графику

**Задание выдал консультант:**

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Ассистент ООД	Немцова О.А.			

**Задание принял к исполнению студент:**

<b>Группа</b>	<b>ФИО</b>	<b>Подпись</b>	<b>Дата</b>
8И6А	Репина Е.Ю.		

## РЕФЕРАТ

Выпускная квалификационная работа бакалавра содержит 107 страниц, 38 рисунков, 33 таблицы, 32 источника, 3 приложения.

Ключевые слова: гибридная модель прогнозирования, штрафной р-сплайн, кластерный анализ, оптимизация параметров, DBScan.

Объект исследования – математический аппарат, предназначенный для решения задачи краткосрочного прогнозирования временных рядов в режиме реального времени.

Целью работы является реализация модели прогнозирования в виде программного приложения, предназначенного для обеспечения одновременно высокого качества прогноза и высокой скорости обработки данных.

Область применения: научно-исследовательская и прикладная сфера обработки данных.

В процессе исследования были рассмотрены несколько вариантов оптимизации параметров модели, отобран наиболее подходящий вариант для получения наилучшего результата. Полученная модель была сопоставлена по ряду критериев (ошибка, производительность) с существующими моделями прогнозирования временных рядов.

На следующем этапе развития проекта планируется расширение функциональных возможностей приложения, повышение производительности модели прогнозирования.

## **ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ**

WF – Windows Forms

MSPE – Mean Squared Percentage Error

MSE – Mean Squared Error

MAPE – Mean Absolute Percentage Error

DBScan - Density-based Spatial clustering of applications with noise

UML - Unified Modeling Language

TU - Theil's U coefficient

POCID - Predicting Changes In Direction

РСС - Рекуррентный Сглаживающий Сплайн

ВР – Временной Ряд

РРВ – Режим Реального Времени

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	14
1. АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ.....	15
1.1. Описание задачи прогнозирования.....	15
1.2. Обзор методов прогнозирования.....	16
1.2.1. Гибридные модели .....	18
1.2.2. Использование сплайн-функций в задаче прогнозирования .....	19
2. РАЗРАБОТКА ГИБРИДНОГО МЕТОДА ПРОГНОЗИРОВАНИЯ .....	21
2.1. Описание математического аппарата .....	21
2.2. Использование метаэвристических алгоритмов для настройки параметров 23	
2.2.1. Генетический алгоритм.....	23
2.2.2. Иммуный алгоритм .....	25
2.2.3. Результаты прогноза при настройке параметров с использованием генетического алгоритма.....	26
2.3. Использование кластерного анализа для настройки параметра $h$ .....	28
2.3.1. Кластеризация временных рядов .....	29
2.3.1.1. Основы кластеризации ВР .....	29
2.3.1.2. Меры сходства в кластеризации временных рядов .....	33
2.3.2. Реализация алгоритмов кластеризации .....	34
2.3.2.1. Алгоритм k-means.....	35
2.3.2.2. Алгоритм DBScan.....	36
3. РАЗРАБОТКА ПРИЛОЖЕНИЯ.....	39
3.1. Описание используемых технологий .....	39
3.2. Классы приложения, их взаимодействие .....	39

3.3. Функциональные требования к программному приложению .....	41
3.4. Интерфейс пользователя.....	43
4. АНАЛИЗ РЕЗУЛЬТАТОВ ПРОГНОЗИРОВАНИЯ.....	45
4.1. Описание конечного алгоритма прогнозирования.....	45
4.2. Описание данных, участвующих в прогнозе.....	45
4.3. Результаты прогнозирования для модельных данных.....	47
4.4. Результаты прогнозирования для реальных данных .....	51
4.4.1. Результаты прогнозирования для данных ПАО «Сбербанк» .....	51
4.4.2. Результаты прогнозирования для данных круизной компании «Carnival Corporation».....	53
4.4.3. Результаты прогнозирования для данных ОАО «Мегафон» .....	54
4.5. Анализ результатов прогнозирования по данным репозитория ICMC-USP Time Series Prediction .....	56
4.5.1. Показатели.....	56
4.5.2. Синтетические данные.....	58
4.5.2.1. Детерминированные временные ряды .....	58
4.5.2.2. Случайные временные ряды.....	60
4.5.2.3. Хаотические временные ряды.....	61
4.5.3. Общие результаты сравнения методов.....	62
5. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ.....	66
5.1. Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения.....	67
5.1.1. Потенциальные потребители результатов исследования.....	67
5.1.2. Анализ конкурентных технических решений .....	68

5.1.3. Технология QuaD.....	70
5.1.4. SWOT-анализ .....	71
5.1.5. Определение возможных альтернатив проведения научных исследований .....	73
5.2. Планирование и формирование бюджета научных исследований.....	74
5.2.1. Структура работ в рамках научного исследования.....	74
5.2.2. Определение трудоемкости выполнения работ .....	75
5.2.3. Разработка графика проведения научного исследования .....	76
5.2.4. Бюджет научно-технического исследования.....	77
5.2.4.1. Расчет материальных затрат научно-технического исследования.....	77
5.2.4.2. Основная заработная плата исполнителей темы.....	78
5.2.4.3. Дополнительная заработная плата исполнителей темы.....	80
5.2.4.4. Отчисления во внебюджетные фонды (страховые отчисления) .....	80
5.2.4.5. Накладные расходы.....	81
5.2.4.6. Формирование бюджета затрат научно-исследовательского проекта.	81
5.3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования.....	82
6. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ.....	85
6.1. Правовые и организационные вопросы обеспечения безопасности .....	86
6.1.1. Специальные правовые нормы трудового законодательства .....	86
6.1.2. Организационные мероприятия при компоновке рабочей зоны.....	87
6.2. Производственная безопасность .....	88
6.2.1. Анализ вредных и опасных факторов .....	89
6.2.1.1. Отклонение показателей микроклимата в помещении .....	89

6.2.1.2. Недостаточная освещенность рабочей зоны .....	91
6.2.1.3. Опасность поражения электрическим током .....	92
6.3. Экологическая безопасность .....	93
6.3.1. Анализ воздействия продукта на окружающую среду .....	93
6.3.2. Решения по обеспечению экологической безопасности .....	94
6.4. Безопасность в чрезвычайных ситуациях .....	94
6.4.1. Перечень возможных ЧС при разработке и эксплуатации .....	94
6.4.2. Разработка действий в результате возникшей ЧС и меры по ликвидации ее последствий .....	95
6.5. Выводы по разделу .....	96
<b>ЗАКЛЮЧЕНИЕ</b> .....	<b>97</b>
<b>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ</b> .....	<b>99</b>
<b>ПРИЛОЖЕНИЕ А</b> .....	<b>103</b>
<b>ПРИЛОЖЕНИЕ Б</b> .....	<b>104</b>
<b>ПРИЛОЖЕНИЕ В</b> .....	<b>103</b>

## ВВЕДЕНИЕ

Понятие "информация" стало одним из ключевых понятий нашего времени. При этом особую ценность представляет информация, полученная при работе с данными, поступающими в режиме реального времени (РРВ). После поступления текущие данные обрабатываются, и система выдает отклик еще до прибытия новой порции данных.

В широком смысле обработкой в режиме реального времени может называться обработка данных за любое время от долей секунды до нескольких минут или даже часов. Что на самом деле здесь важно - это задержка между получением данных и созданием информации. Системы в реальном масштабе времени используются в службах управления воздушным движением, в автоматических банках данных, в автопилотах летательных аппаратов и во многих военных прикладных программах.

Обработанные данные используются в различных задачах интерпретации: фильтрации, прогнозирования, дифференцирования и др. При этом все данные содержат шум. Поэтому важную роль играет аппарат фильтрации и сглаживания данных, при помощи которого значительно повышается точность задачи интерпретации.

Одной из форм представления цифровых данных являются временные ряды. Сглаживание временных рядов лежит в основе многих прикладных задач. Применение сплайна позволяет получить искомое гладкое решение.

# 1. АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

## 1.1. Описание задачи прогнозирования

В настоящее время задача прогнозирования является актуальной практически для всех предметных областей и сфер деятельности. Частным случаем задачи прогнозирования является прогнозирование временных рядов, как одной из популярных форм представления информации. Значения какого-либо параметра предметной области фиксируются через определенные временные интервалы и образуют временной ряд. В таком случае прогноз представляет собой предсказание нового значения еще до момента его получения на основе имеющихся данных.

Разработано множество моделей для решения задачи прогнозирования временного ряда, среди которых наибольшую применимость имеют авторегрессионные и нейросетевые модели [11]. Большую ценность представляют модели, способные работать с прогнозом в режиме РРВ. С помощью данных, полученных при таком прогнозе, можно оперативно и объективно отражать постоянно обновляемую картину любого процесса. Подобная задача относится к краткосрочному прогнозированию временных рядов. Краткосрочный прогноз представляет собой прогноз на 1- 3 значения вперед.

Краткосрочное прогнозирование выступает как элемент корректировки среднесрочных и долгосрочных прогнозов и используется в узких сферах требующих оперативного управления [16]. Так, например, в экономике к краткосрочному прогнозированию прибегают для определения факторов экономического роста, конкурентоспособности, промышленной политики, моделирования и прогнозирования развития реального сектора. Еще один пример использования краткосрочного прогнозирования – прогнозирование ЧС. Краткосрочные прогнозы используются для ориентировочного определения времени возникновения ЧС. Для составления прогнозов

используются различные статистические данные, а также сведения о некоторых физических и химических характеристиках окружающих природных сред.

## **1.2. Обзор методов прогнозирования**

В наше время насчитывается большое количество классов моделей прогнозирования (более сотни). Большинство из них основываются на других моделях, следовательно, число неповторимых классов моделей гораздо меньше [21]. Часть моделей и соответствующих методов относится к отдельным процедурам прогнозирования. Часть методов представляет набор отдельных приемов, отличающихся от базовых или друг от друга количеством частных приемов и последовательностью их применения.

Все методы прогнозирования можно разделить на две группы: интуитивные и формализованные.

Интуитивное прогнозирование применяется тогда, когда объект прогнозирования либо слишком прост, либо, напротив, настолько сложен, что аналитически учесть влияние внешних факторов невозможно. Интуитивные методы прогнозирования не предполагают разработку моделей прогнозирования и отражают индивидуальные суждения специалистов (экспертов) относительно перспектив развития процесса. Интуитивные методы основаны на мобилизации профессионального опыта и интуиции. Такие методы используются для анализа процессов, развитие которых либо полностью, либо частично не поддается математической формализации, то есть для которых трудно разработать адекватную модель. К подобным методам относятся следующие методы: экспертных оценок, исторических аналогий, предвидения по образцу. Формализованные методы рассматривают модели прогнозирования. Классификация методов и моделей прогнозирования представлена на рисунке 1.

В статистических моделях функциональная зависимость между будущими и фактическими значениями временного ряда, а также внешними

факторами задана аналитически. К статистическим моделям относятся следующие группы [21]:

1. Регрессионные модели (линейная регрессия, нелинейная регрессия);
2. Авторегрессионные модели (ARIMAX, GARCH, ARDLN);
3. Модель экспоненциального сглаживания;
4. Модель по выборке максимального подобия и т.д.

В структурных моделях функциональная зависимость между будущими и фактическими значениями временного ряда, а также внешними факторами задана структурно. К структурным моделям относятся следующие группы [21]:

1. Нейросетевые модели;
2. Модели на базе цепей Маркова;
3. Модели на базе классификационно-регрессионных деревьев и т.д.

Помимо этого для узкоспециализированных задач иногда применяются особые модели прогнозирования.



Рисунок 1 – Классификация методов и моделей прогнозирования

### 1.2.1. Гибридные модели

Одной из популярных современных тенденций в области создания моделей прогнозирования является создание комбинированных моделей и методов. Подобный подход дает возможность компенсировать недостатки одних моделей при помощи других, и направлен на повышение точности прогнозирования, как одного из главных критериев эффективности модели [15].

В некоторых работах процесс прогнозирования разделяют на два этапа [15,16]. На первом этапе с помощью какого-либо инструмента весь ряд разделяют на группы (кластеры) временного ряда. На следующем этапе для каждой группы строится отдельная модель прогнозирования. При подобном подходе удастся повысить точность прогнозирования временных рядов.

При обзоре гибридных моделей прогнозирования были выделены следующие популярные связки:

- нейронные сети + ARIMA (ARIMAX)
- нейронные сети + регрессия
- нейронные сети + нечеткая логика (возможно добавление GA)
- регрессия + нечеткая логика

В большинстве комбинаций модели на основе нейронных сетей применяются для решения задачи кластеризации, а далее для каждого кластера строится отдельная модель прогнозирования на основе ARIMA, GA, нечеткой логики и др. Также применяются другие методы кластеризации.

Применение гибридных моделей является направлением, которое при корректном подходе позволяет повысить точность прогнозирования. Недостаток таких моделей - сложность и ресурсоемкость их разработки: нужно разработать модели таким образом, чтобы компенсировать недостатки каждой из них, не потеряв достоинств.

### **1.2.2. Использование сплайн-функций в задаче прогнозирования**

Один из отличительных подходов к задаче прогнозирования представляет собой использование аппроксимирующих кусочно-полиномиальных функций, заменяющих регрессионные соотношения при моделировании и анализе процессов. С помощью подобного аппарата выполняется фильтрация и сглаживание данных с шумом, что значительно повышает точность задачи интерпретации [14].

При этом используются различные виды сплайнов: регрессионные сплайны, B-сплайны, штрафные сплайны, сглаживающие сплайны и т.д.

Сравнение результатов использования различных сплайн-функций в задаче прогнозирования [18] показывает, что результат, полученный с помощью штрафных сплайнов, незначительно отличается от результата, полученного с помощью сглаживающих сплайнов (кубических), но все таки

превосходит его (согласно значениям критерия оценки MSE, MAE и MAPE для временного ряда).

## 2. РАЗРАБОТКА ГИБРИДНОГО МЕТОДА ПРОГНОЗИРОВАНИЯ

В данной работе будет рассмотрен следующий комбинированный подход к прогнозированию временного ряда:

- На первом этапе будет производиться настройка параметров сглаживающего сплайна с помощью кластеризации;
- На втором этапе будет построен штрафной P-сплайн с оптимальными параметрами (полученными на предыдущем этапе) и точкой сопряжения звеньев  $m$  ( $m=h+1$ ), т.е. прогноз на одну точку вперед

Пересчет всего набора данных для настройки параметров при поступлении каждой новой точки в режиме реального времени является слишком ресурсоемким и излишним процессом. Для адаптации данной модели прогнозирования к режиму реального времени необходимо выделить некоторое число последних кластеров (LC), данные которых совместно с новоприбывшим значением будут являться начальным набором данных для повторного запуска работы алгоритма кластеризации.

### 2.1. Описание математического аппарата

Одной из форм представления цифровых данных являются временные ряды. Сглаживание временных рядов лежит в основе многих прикладных задач.

Применение сплайна позволяет получить искомое гладкое решение [9]. Сглаживающий сплайн  $S(t)$  основан на оптимизации специального вида функционала и представлен для РРВ следующей формулой:

$$J(S) = (1 - \rho)(h\Delta t)^2 \int_{t_0^i}^{t_h^i} [S''(t)]^2 dt + \rho \sum_{j=0}^h [S(t_j^i) - y(t_j^i)]^2 \quad (1)$$

где  $\rho \in [0, 1]$  — весовой коэффициент, устанавливающий баланс между сглаживающими и интерполяционными свойствами сплайна  $S(t)$ ;

$\Delta t$  — интервал дискретизации наблюдаемого процесса;

$h$  — количество измерений внутри  $i$ -го звена сплайна.

Для РРВ используется рекуррентная форма расчета коэффициентов сплайна для каждого  $i$ -го звена на основании  $h$  отсчетов:

$$S_i(m) = a_0^i + a_1^i m + a_2^i m^2 + a_3^i m^3, -q \leq m \leq h - q \quad (2)$$

где  $q$  – номер отсчета  $t_j^i$  внутри  $i$ -го звена ( $j = \overline{0, h}; q = \overline{0, h - 1}$ ), в котором сопряжены непрерывные производные сплайна;  $m = t_j^i - t_0^i$  номер отсчета  $t_j^i$  внутри  $i$ -го звена, в котором сплайн вычисляется как  $m = \overline{-q, h - q}$ :  $h = (t_h^i - t_0^i)/\Delta t$  – количество отсчетов в  $i$ -м звене сплайна.

Разрывные коэффициенты находятся из условия минимизации экстремального функционала (1) в зависимости от дефекта сплайна  $d$ .

В зависимости от соотношения моментов сопряжения  $q \in [0, h - 1]$  и вычисления  $m \in [1, h]$   $i$ -го сплайна выделяют несколько режимов функционирования РСС (рис.2): последовательный режим функционирования сплайна, текущий режим функционирования и режим многократного сглаживания [9].

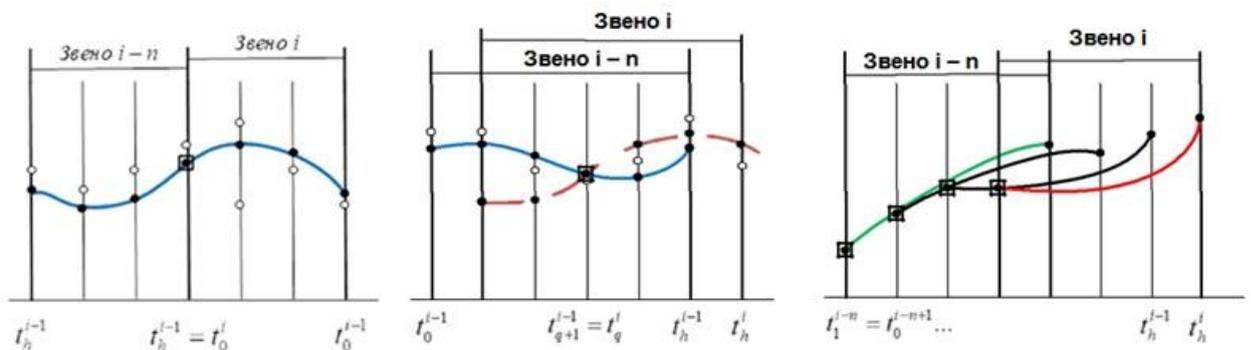


Рисунок 2 – Режимы функционирования РСС (последовательный, текущий, режим многократного сглаживания)

Для вычисления прогноза в режиме РРВ используется текущий режим функционирования сплайна – звено сплайна вычисляется при поступлении нового значения с использованием  $h-1$  предыдущих значений.

В таком режиме формулы для оценки коэффициентов принимают следующий вид (при  $q = 0$  и дефект  $d = 2$ ):

$$a_0^i = a_0^{i-1} + a_1^i h + a_2^i h^2 + a_3^i h^3$$

$$a_1^i = a_1^{i-1} + 2a_2^i h + 3a_3^i h^2 \quad (3)$$

$$a_2^i = \frac{\rho(F_1^i C - F_2^i A)}{BC - A^2}$$

$$a_3^i = \frac{\rho(F_2^i B - F_1^i A)}{BC - A^2}$$

При этом:

$$A = 6(1 - \rho)h^4 + \rho H_5$$

$$B = 4(1 - \rho)h^3 + \rho H_4$$

$$C = 12(1 - \rho)h^5 + \rho H_6 \quad (4)$$

$$F_1^i = \sum_{k=0}^h y(t_k^i) k^2 - a_0^i H_2 - a_1^i H_3$$

$$F_2^i = \sum_{k=0}^h y(t_k^i) k^3 - a_0^i H_3 - a_1^i H_4$$

$$H_n = \sum_{k=0}^h k^n$$

## 2.2. Использование метаэвристических алгоритмов для настройки параметров

В качестве инструмента настройки параметров сплайна можно выделить следующие метаэвристические алгоритмы: генетический алгоритм, иммунный алгоритм.

### 2.2.1. Генетический алгоритм

Генетический алгоритм – это эвристический алгоритм поиска, используемый для решения задач оптимизации и моделирования путём

случайного подбора, комбинирования и вариации искомых параметров с использованием механизмов, аналогичных естественному отбору в природе.

Данный метод является методом перебора решений для тех задач, в которых невозможно найти решение с помощью математических формул. Однако простой перебор решений в сложной многомерной задаче – это бесконечно долго. Поэтому генетический алгоритм перебирает не все решения, а только лучшие. Алгоритм берёт группу решений и ищет среди них наиболее подходящие. В первую очередь генерируется начальная популяция. Далее происходит процесс скрещивания (размножения). Основной принцип размножения — потомок похож на своих родителей. Т.е. должен быть задан какой-то механизм наследования. Затем происходит небольшое изменение в полученных особях (процесс мутации) – получает новые решения, среди которых снова отбирает лучшие, а худшие отбрасывает. Таким образом, на каждом шаге работы алгоритм отбирает наиболее подходящие решения (проводит селекцию), считая, что они на следующем шаге дадут ещё более лучшие решения (эволюционируют).

Этих трех механизмов чаще всего недостаточно, чтобы минимизировать функцию. Так популяция вырождается — рано или поздно локальный минимум забивает своим значением всю популяцию. Когда такое происходит, проводят процесс, называемый встряской (в природе аналогии — глобальные катаклизмы), когда уничтожается почти вся популяция, и добавляются новые (случайные) особи. Принцип работы генетического алгоритма представлен на рисунке 3.

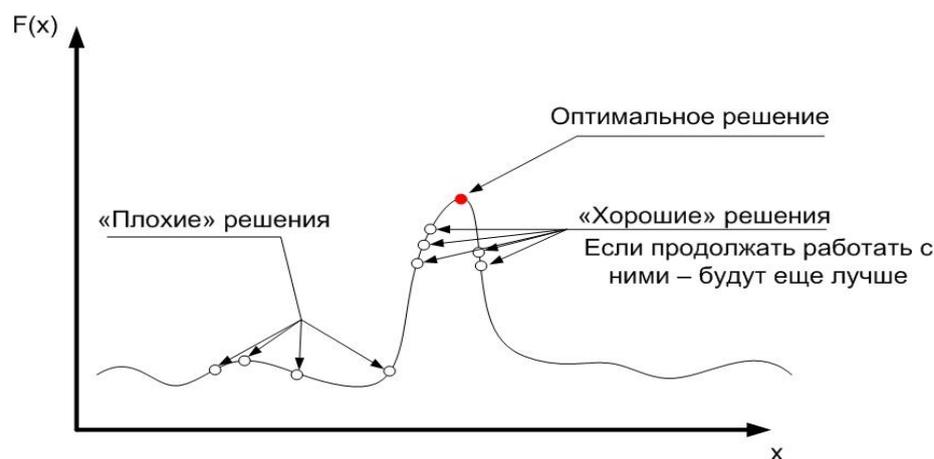


Рисунок 3 – Интерпретация генетического алгоритма

### 2.2.2. Иммунный алгоритм

В некотором роде иммунные системы можно считать наследниками генетических алгоритмов и нейронных сетей, обладающими определенной спецификой. Лимфоциты – клетки, из которых состоит иммунная система. Если пользоваться терминами генетических алгоритмов, лимфоцит – это просто особь. Аффинность – мера того, насколько хорошо данный лимфоцит реагирует на какой-то антиген (вирус). В биологии это определяется различными химическими связями и реакциями, в данном алгоритме это просто значение целевой функции, которая представляет собой евклидово расстояние между точными заданными значениями функции и значениями полученной функции в тех же точках.

Первоначально создается некоторое множество случайным образом сгенерированных лимфоцитов (в организме этим занимается костный мозг). Затем на протяжении всего цикла функционирования системы из текущего множества лимфоцитов выбираются лучшие, к ним применяются различные операции гипермутации (для создания лучше приспособленных клеток – эволюция в действии). Затем из старых и вновь полученных клеток выбирается новое текущее множество, и данный шаг повторяется заново до тех пор, пока не будет найдено решение с достаточной точностью, или мы проведем максимально допустимое число итераций.

### 2.2.3. Результаты прогноза при настройке параметров с использованием генетического алгоритма

В качестве инструмента настройки параметров сплайна был выбран генетический алгоритм, т.к. данный алгоритм превосходит подобные алгоритмы в быстрой работе. С помощью данного алгоритма были получены оптимальные значения для основных параметров ( $h$  и  $\rho$ ).

Результаты проверки полученного метода на модельных данных представлены в таблице 1 и на рисунках 4, 5.

Таблица 1 – Результаты прогноза модельных данных

Функция №1	
$h_{opt}$	4
$\rho_{opt}$	0.37
$MSPE_{min}$	3.22
Функция №2	
$h_{opt}$	3
$\rho_{opt}$	0.55
$MSPE_{min}$	5.31

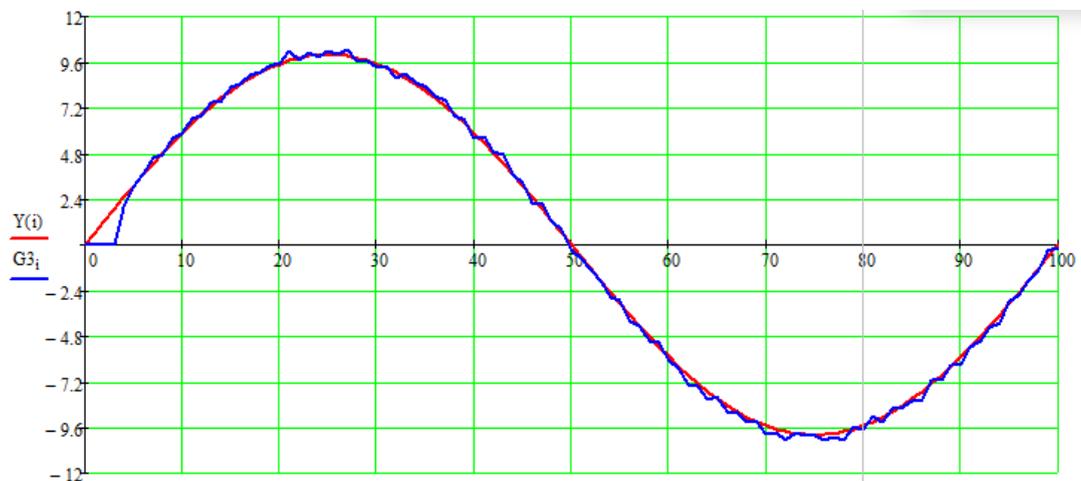


Рисунок 4 – Результат прогнозирования для функции №1 при шуме в 5 %

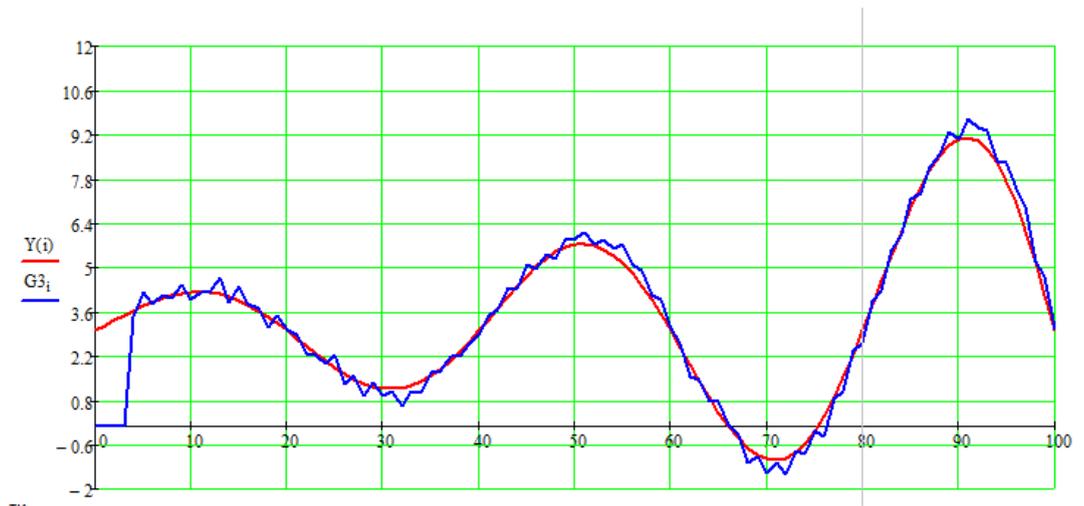


Рисунок 5 – Результат прогнозирования для функции №2 при шуме в 5 %

Для апробации полученного метода на основе реальных данных были использованы следующие наборы данных (за несколько последних лет):

- Значения цены закрытия акций компании «Мегафон»;
- Значения цены закрытия акций банка «Сбербанк».

Полученные результаты представлены в таблице 2 и на рисунках 6, 7.

Таблица 2 – Результаты прогноза реальных данных

«Сбербанк»	
$h_{opt}$	10
$\rho_{opt}$	0.23
$MSPE_{min}$	10.13
«Мегафон»	
$h_{opt}$	9
$\rho_{opt}$	0.61
$MSPE_{min}$	12.67

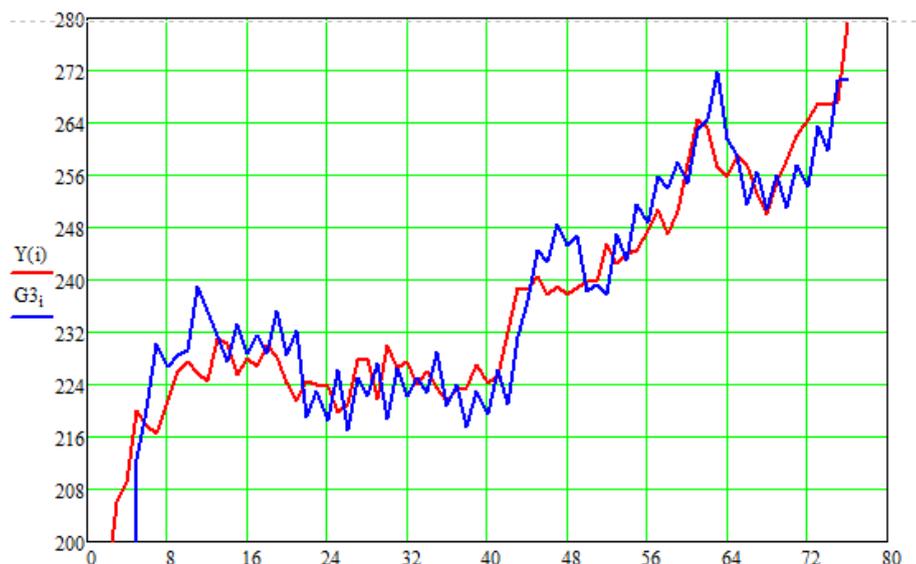


Рисунок 6 – Результат прогнозирования акций компании «Сбербанк»

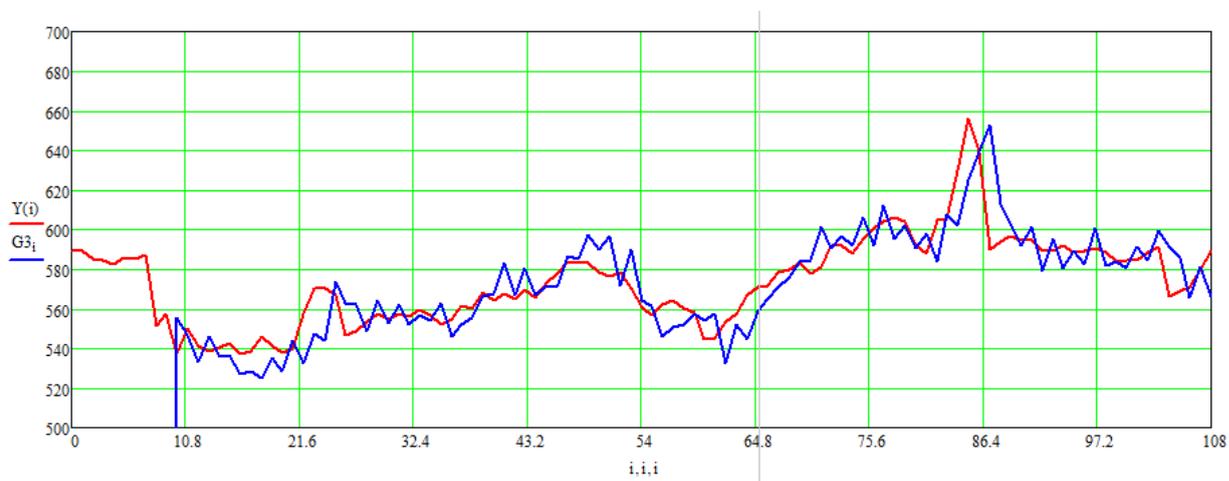


Рисунок 7 – Результат прогнозирования акций компании «Мегафон»

Анализ полученных результатов показал, что данный метод показывает приемлемые результаты на несложных модельных данных – даже на данных с шумом в 20% погрешность не превышает значения в 6%. Для реальных данных результат прогнозирования оказался неудовлетворительным - даже при небольшом шуме погрешность превышает значение 10%.

### 2.3. Использование кластерного анализа для настройки параметра $h$

Все параметры сплайна оказывают определенное влияние на его свойства. Наиболее значимый вклад в точность аппроксимации вносит

параметр  $h$ . Современные методы оптимизации, включая биоинспирированные и другие метаэвристики, позволяют получить только постоянное значение параметра  $h$  для всего интервала наблюдения. Однако это приводит к повышению погрешности в конечном результате.

Одним из возможных способов разбиения временного ряда на сегменты с переменным параметром числом измерений  $h$  является кластерный анализ.

### **2.3.1. Кластеризация временных рядов**

#### **2.3.1.1. Основы кластеризации ВР**

Кластеризацией называют разбиение множества объектов на группы (кластеры).

Все алгоритмы кластеризации принято делить на иерархические и неиерархические (по данным, получаемым на выходе). Иерархические алгоритмы делятся на агломеративные и дивизимные.

Агломеративные алгоритмы - это алгоритмы, которые начинают свое выполнение с того, что каждый объект заносит в свой собственный кластер и по мере выполнения объединяют кластеры, до тех пор, пока в конце не получает один кластер, включающий в себя все объекты набора.

Дивизимные алгоритмы, напротив, сначала относят все объекты в один кластер и затем разделяют этот кластер до тех пор, пока каждый объект не окажется в своем собственном кластере. Неиерархические алгоритмы кластеризации можно разделить по принципам кластеризации: итеративные, плотностные, сетевые, модельные и концептуальные.

Итеративные алгоритмы кластеризации – предполагают пошаговое перераспределение объектов между классами. Одним из популярных представителей данного семейства алгоритмов является алгоритм k-means. Основная идея данного алгоритма - пошаговая минимизация расстояний между объектами в кластерах до тех пор, пока это возможно. Главным недостатком

алгоритма, применительно к РРВ, является необходимость заранее задавать желаемое количество кластеров, что недопустимо в РРВ.

Плотностные алгоритмы кластеризации - алгоритмы, определяющие кластер как группу объектов, расположенных кучно, т.е. так, чтобы в  $\epsilon$ -окрестности точки находилось минимально заданное число других объектов (соседей). Представителем этого класса алгоритмов кластеризации является алгоритмы DBScan, OPTICS.

Модельные алгоритмы подразумевают, существует некоторая модель кластера (его структуры), и стремится найти и максимизировать сходства между этой моделью и имеющимися данными, то есть выделить в данных такие модели, которые и будут кластерами (пример: Алгоритм EM) [22].

Одним из представителей концептуальных алгоритмов является алгоритм COBWEB. В алгоритме реализовано вероятностное представление категорий. Принадлежность категории определяется не набором значений каждого свойства объекта, а вероятностью появления значения. Для каждой категории в иерархии определены вероятности вхождения всех значений каждого свойства. При предъявлении нового экземпляра система COBWEB оценивает качество отнесения этого примера к существующей категории и модификации иерархии категорий в соответствии с новым представителем.

Алгоритм WaveCluster является представителем группы сетевых алгоритмов. Алгоритм рассматривает всю совокупность данных как сигнал в  $N$ -мерном пространстве атрибутов и пытается на основании вейвлет-преобразования выделить в этом сигнале поддиапазоны частот, в которых связанные компоненты и будут кластерами.

Несмотря на то, что временной ряд состоит из большого числа точек данных, его можно рассматривать как один объект. Данные временных рядов представляют большой интерес из-за их повсеместного использования. Именно поэтому существует большое количество работ, направленных на изучение временных рядов. Данные работы преследуют различные цели: подбор

подпоследовательности, обнаружение аномалий, обнаружение мотива, индексация, кластеризация, классификация, визуализация, сегментация, выявление закономерностей, анализ тенденций, обобщение и прогнозирование [22].

За последнее десятилетие в области кластеризации временных рядов произошли значительные изменения, вызванные появлением таких концепций, как большие данные и облачные вычисления, которые экспоненциально увеличили размер наборов данных. Кластеризация таких сложных объектов особенно полезна, поскольку она приводит к обнаружению интересных закономерностей в наборах данных временных рядов. Для того чтобы обосновать важность задачи кластеризации применительно к временным рядам, приведем цели данной кластеризации:

1. Данные временных рядов содержат ценную информацию, которую необходимо вычлениить из общей массы информации с помощью обнаружения шаблонов. Кластеризация - это

распространенное решение, выполняемое для обнаружения этих шаблонов в наборах данных временных рядов.

2. Данные временных рядов имеют настолько большие объемы, что не могут быть качественно обработаны человеком. Для решения этой проблемы данные должны быть сначала хорошо структурированы, а только потом переданы на обработку человеку. Для структурирования данных также используется кластеризация.

3. Кластеризация временных рядов используется в качестве исследовательского метода, а также в качестве промежуточного этапа более сложных алгоритмов извлечения данных, таких как: обнаружение правил, обнаружения аномалий, индексация.

4. С помощью кластеризации может быть реализована визуализация данные временных рядов (представление структур кластеров временных рядов

в виде визуальных изображений), которая позволяет быстро понять структуру данных, кластеров, аномалий и других закономерностей в наборах данных.

Кластеризация временных рядов является сложной задачей и имеет ряд проблем. Во-первых, данные временных рядов часто намного больше, чем объем памяти, и, следовательно, они хранятся на дисках. Это приводит к экспоненциальному снижению скорости процесса кластеризации. Во-вторых, данные временных рядов часто являются многомерными, что затрудняет обработку этих данных для многих алгоритмов кластеризации, а также замедляет процесс кластеризации. Третья проблема – сложность нахождения мер сходства, которые используются для создания кластеров.

Итак, кластеризация временных рядов используется для обнаружения некоторых закономерностей. Нахождение кластеров в временных рядах может быть использовано для решения следующих задач:

1. Обнаружение аномалий, новизны или диссонанса. Обнаружение аномалий - это методы обнаружения необычных и неожиданных паттернов, которые неожиданно происходят в наборах данных.

Например, в базах данных датчиков, кластеризация временных рядов, которые производятся по показаниям датчиков мобильного робота для обнаружения событий.

2. Распознавание динамических изменений во временных рядах: обнаружение корреляции между временными рядами. Например, в финансовых базах данных его можно использовать для поиска компаний с аналогичным движением цены акций.

3. Предсказание и рекомендация: гибридный метод, сочетающий кластеризацию и приближение функций на кластер, может помочь пользователю прогнозировать и рекомендовать. Например, в научных базах данных он может решать такие проблемы, как поиск моделей солнечного магнитного ветра для прогнозирования сегодняшней картины.

4. Обнаружение паттернов: обнаружение интересных паттернов в базах данных. Например, в маркетинговой базе данных могут быть обнаружены различные ежедневные схемы продаж определенного товара в магазине.

### **2.3.1.2. Меры сходства в кластеризации временных рядов**

Кластеризация временных рядов в значительной степени зависит от измерения расстояния. Существуют различные меры, которые могут применяться для измерения расстояния между временными рядами. В традиционной кластеризации расстояние между статическими объектами точно основано на совпадении, но при кластеризации временных рядов расстояние вычисляется приблизительно. В частности, для сравнения временных рядов с нерегулярными интервалами и длиной выборки очень важно адекватно определить сходство временных рядов. Существуют различные меры расстояния, предназначенные для определения сходства между временными рядами. Одним из самых простых способов вычисления расстояния между двумя временными рядами является рассмотрение их как одномерных временных рядов, а затем вычисление измерения расстояния по всем временным точкам. Подходы кластеризации могут быть классифицированы на две категории на основе продолжительности временных рядов: «уровень формы» и «уровень структуры». «Уровень формы» обычно используется для измерения сходства в кластеризации коротких временных рядов путем сравнения их локальных паттернов, тогда как «уровень структуры» измеряет сходство, основанное на глобальной и высокоуровневой структуре. Выбор меры расстояния является спорным вопросом, но есть некоторые общие замечания:

- Из всех методов выбора расстояния наиболее эффективными и точными являются те, которые основаны на динамическом программировании (DP). Эти методы имеют низкую скорость и требуют тщательной настройки параметров, чтобы быть эффективными и действенными.

· Существует проблема несовместимости метрики расстояния с методом представления.

· Самые распространенные методы - евклидово расстояние и DTW. Евклидово расстояние удивительно конкурентоспособно, однако DTW также имеет свои преимущества в измерениях подобия.

Мера сходства, используемая в реализованном алгоритме - евклидово расстояние.

### 2.3.2. Реализация алгоритмов кластеризации

В ходе выполнения работы были реализованы и опробованы два алгоритма кластеризации: k-means и DBScan. Оба алгоритма нуждались в модификации для работы с прогнозом в режиме РРВ. Оба алгоритма были опробованы на тестовых данных для следующих функций:

Функция №1:

$$f_1(t) = 10 \sin\left(\frac{2\pi t}{100}\right), \quad (5)$$

Функция №2:

$$f_2(t) = \sin\left(\frac{\pi t}{20}\right)e^{0,02t} + 3. \quad (6)$$

После чего было произведено сравнение результатов работы алгоритмов. Результаты сравнения представлены в таблице 3.

Таблица 3 – Сравнение алгоритмов кластеризации

	Плюсы	Минусы
<b>K-means</b>	<ul style="list-style-type: none"><li>• Нет выбросов</li><li>• Непересекающиеся кластеры</li></ul>	<ul style="list-style-type: none"><li>• Необходимость задания числа кластеров до начала работы алгоритма</li><li>• Низкая производительность</li></ul>

<b>DBScan</b>	<ul style="list-style-type: none"> <li>• Хорошо выделяет момент смены тенденций в данных</li> <li>• Приемлемая производительность</li> <li>• Приспособлен для работы на данных с шумом</li> </ul>	<ul style="list-style-type: none"> <li>• Пересекающиеся кластеры</li> <li>• Выделяет выбросы в данных</li> </ul>
---------------	---	--

В конечном итоге был выбран алгоритм DBScan. Данный алгоритм был модифицирован и настроен для работы с данными в режиме PPV.

### 2.3.2.1. Алгоритм k-means

Для начала работы с кластеризацией данных был выбран алгоритм k-means. Наиболее простой, но в то же время достаточно неточный метод кластеризации в классической реализации. Он разбивает множество элементов векторного пространства на заранее известное число кластеров  $k$ . Действие алгоритма таково, что он стремится минимизировать среднеквадратичное отклонение на точках каждого кластера. Основная идея заключается в том, что на каждой итерации перевычисляется центр масс для каждого кластера, полученного на предыдущем шаге, затем векторы разбиваются на кластеры вновь в соответствии с тем, какой из новых центров оказался ближе по выбранной метрике. Алгоритм завершается, когда на какой-то итерации не происходит изменения кластеров. Алгоритм был реализован для некоторого конечного числа данных и заданном количестве кластеров.

Одним из минусов алгоритма является необходимость точного определения количества кластеров еще до начала работы, что невозможно при работе с данными в режиме реального времени. Еще один минус – низкая производительность.

Результат работы алгоритма на модельных данных представлен на рисунках 8, 9.

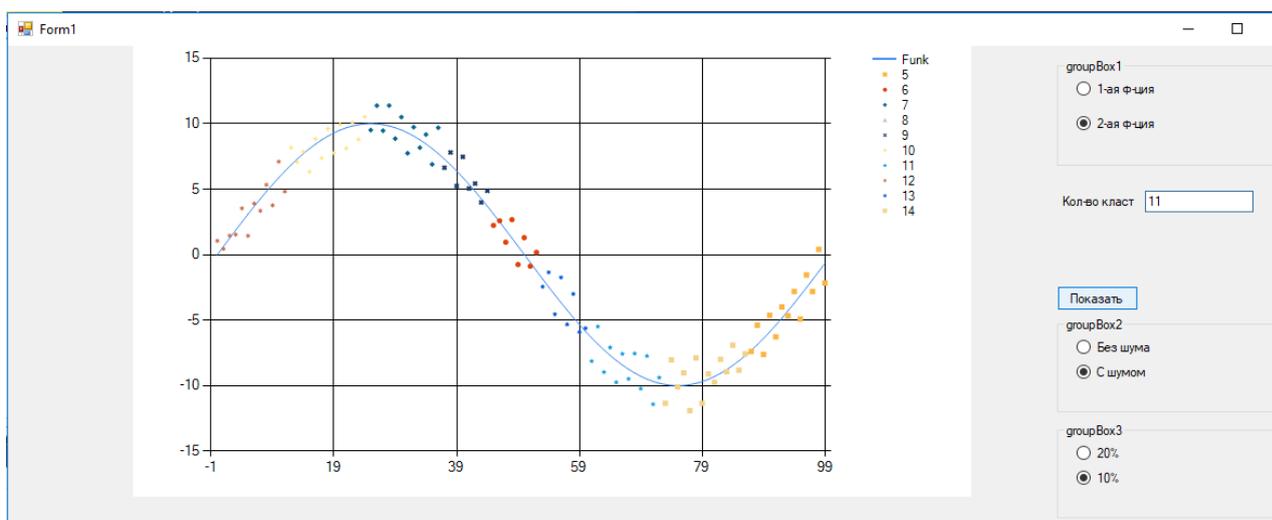


Рисунок 8 – Результаты работы алгоритма k-means для функции №1 с шумом в 10%

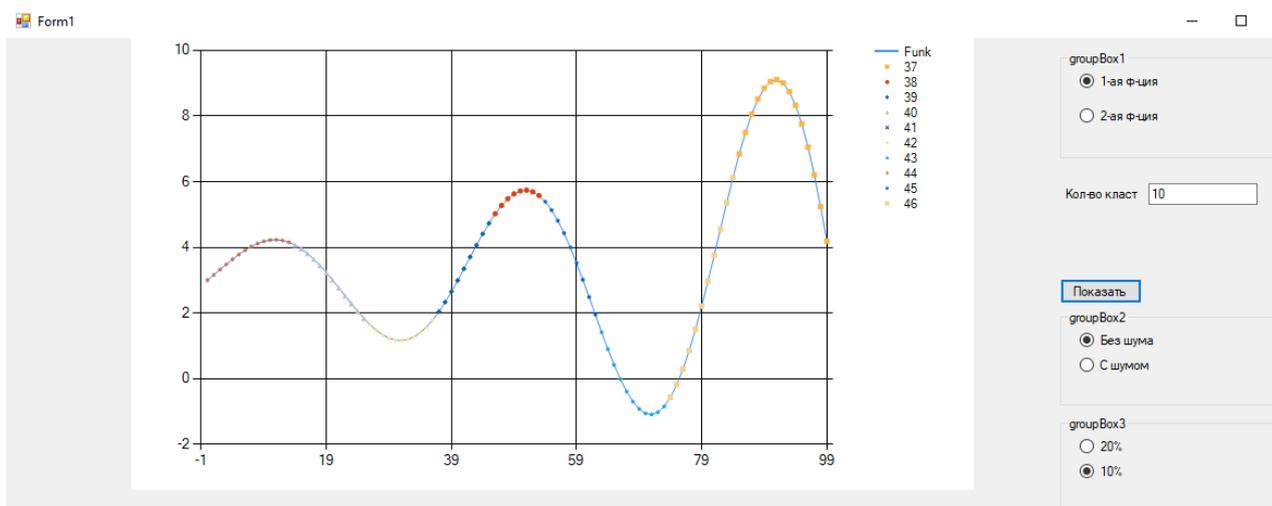


Рисунок 9 – Результаты работы алгоритма k-means для функции №2 с шумом в 0%

Данную проблему можно решить с помощью использования одного из методов определения оптимального количества кластеров внутри конечного набора данных.

### 2.3.2.2. Алгоритм DBScan

Алгоритм DBScan (Density-based spatial clustering of applications with noise) допускает кластеризацию пространственных данных в присутствии

шума. Алгоритм был предложен в 1996 году М. Эстером и его коллегами для разбиения данных на кластеры произвольной формы [10]. Для работы алгоритма используются два входных параметра:  $\epsilon$ -окрестность, в которой будет требоваться наличие минимального количества объектов  $Minpts$ . Данный параметр ограничен минимально возможным числом измерений для построения звена сплайна, и, следовательно, не может быть уменьшен. Увеличение этого параметра в задачах РРВ нелогично.

Результат работы алгоритма на модельных данных представлен на рисунках 10, 11.

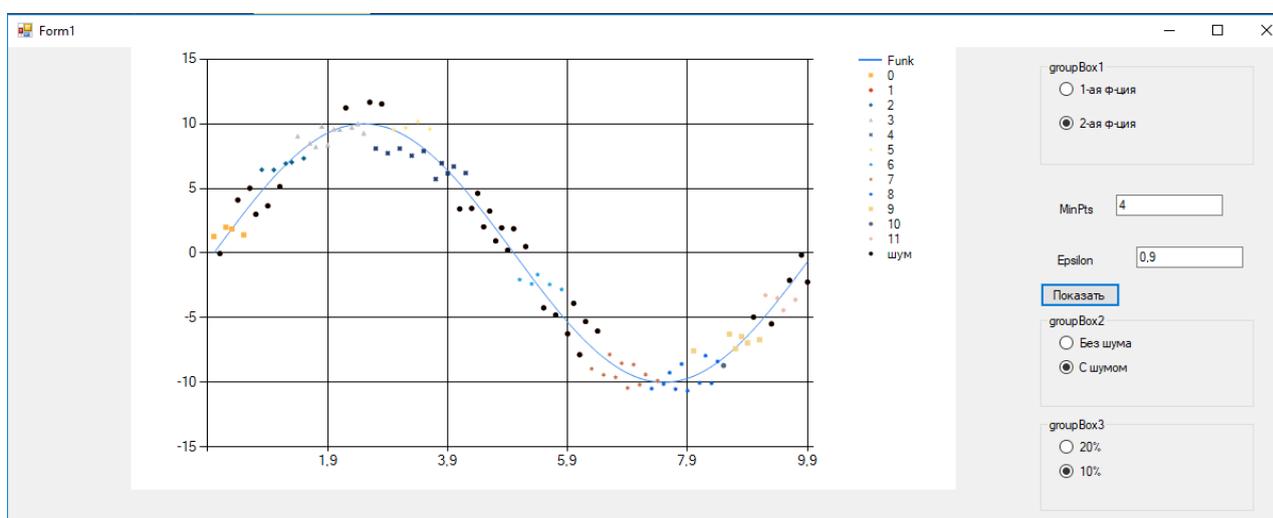


Рисунок 10 – Результаты работы алгоритма DBScan для функции №2 с шумом в 10%

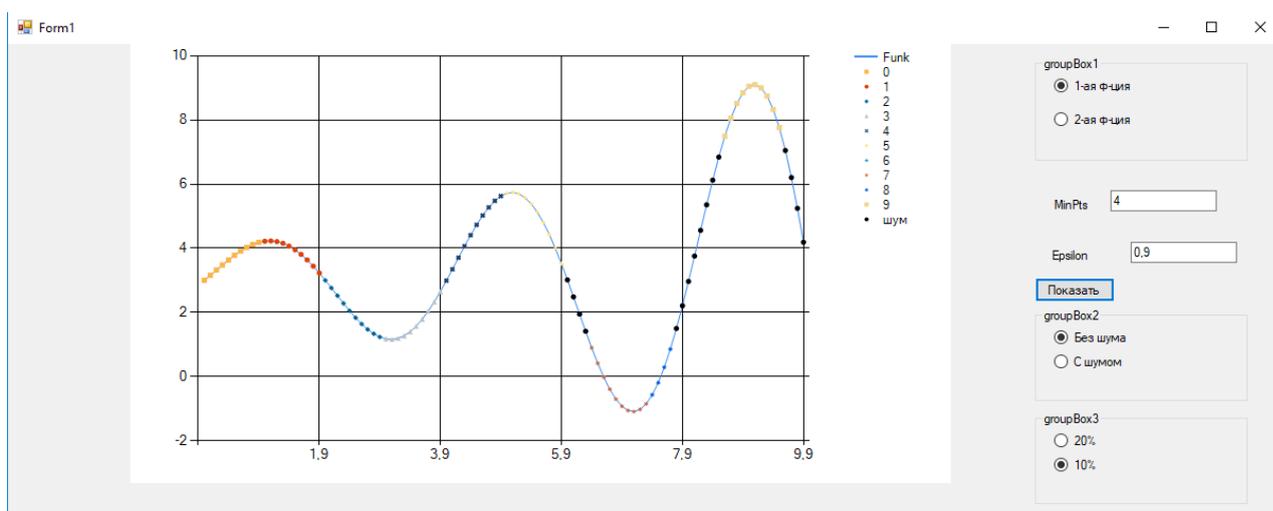


Рисунок 11 – Результаты работы алгоритма DBScan для функции №2 с шумом в 0%

Классическая реализация алгоритма является неприменимой для поставленной задачи прогнозирования по ряду причин. Алгоритм не реализует ограничение количества элементов сверху, что приводит к неприемлемо большому для точного сглаживания кластерам. Также, алгоритм предусматривает отбрасывание "шума", что недопустимо для задач прогнозирования, т.к. это может привести к потере важной информации.

Данный алгоритм был модифицирован под решаемую задачу: задание верхней границы количества элементов в кластере, решение проблемы перекрытия кластеров, выделение последовательных точек «шума» в отдельные кластеры, создание правил для применения алгоритма в режиме реального времени.

Для того чтобы прогноз можно было получить в режиме реального времени, необходимо определиться с тем, как будет работать алгоритм при поступлении новых данных. Пересчет всего набора данных при поступлении каждой новой точки является ресурсоемким и излишним процессом, поэтому было принято решение работать только с последним кластером. Данный кластер и только что поступившая точка являются входными данными для работы алгоритма кластеризации.

### **3. РАЗРАБОТКА ПРИЛОЖЕНИЯ**

#### **3.1. Описание используемых технологий**

Для реализации был выбран язык C#. C# — объектно-ориентированный язык программирования. Язык имеет статическую типизацию, поддерживает полиморфизм, перегрузку операторов (в том числе операторов явного и неявного приведения типа), делегаты, атрибуты, события, свойства, обобщённые типы и методы, итераторы, анонимные функции с поддержкой замыканий, LINQ и т.д.

В качестве среды разработки используется Microsoft Visual Studio 2019. Для создания графических интерфейсов с помощью платформы .NET применяются разные технологии - Window Forms, WPF и т.д.

Для реализации приложения была выбрана платформа Window Forms (WF). Данная платформа является наиболее простой и удобной, а также имеет оптимальный функционал для решения поставленной задачи.

Windows Forms позволяет разрабатывать интеллектуальные клиенты [19]. Интеллектуальный клиент — это приложение с полнофункциональным графическим интерфейсом, простое в развертывании и обновлении, способное работать при наличии или отсутствии подключения к Интернету и использующее более безопасный доступ к ресурсам на локальном компьютере по сравнению с традиционными приложениями Windows.

#### **3.2. Классы приложения, их взаимодействие**

Обработка взаимодействия приложения и пользователя реализована путем добавления обработчиков стандартных элементов форм. Основная логика программы была вынесена в отдельные классы и методы. Диаграмма классов и их описание представлены далее:

Data() – в данном классе осуществляется большая часть работы с данными: представление данных, их обработка, визуализация и т.д. На модельные данные в этом классе накладывается шум в 10%.

MyCustomDatasetItem() – в классе реализовано двумерное представление данных.

NewPoint() – в данном классе реализуется пошаговое добавление новых значений при обращении пользователя (имитация работы поступления новых данных в режиме РРВ). Здесь же реализовано объединение поступившего значения с последним кластером, проверка данной выборки на полноту.

DbscanAlgorithm() – в данном классе реализован алгоритм кластеризации: разбиение данных на кластеры, проверка границ количества данных в кластерах, представление кластеров для других классов.

DbscanPoint() – в классе реализовано представление данных для дальнейшей работы алгоритма кластеризации (каждому значению задается состояние и номер кластера).

Forecasting() – здесь происходит вычисление прогнозного значения по входным данным: пересчет необходимых коэффициентов, вспомогательных функций. В этом классе происходит пополнение данных прогноза и сглаживания, обработка и отображение ошибок.

Error() – класс содержит методы для вычисления ошибок прогнозирования: среднеквадратическая ошибка прогнозирования по всем полученным значениям и абсолютная ошибка для каждого нового значения.

ShowChart() – класс, предназначенный для визуализации получаемых данных и прогнозных значений.

ShowClusters() - класс, предназначенный для визуализации разбиения данных на кластеры.

Взаимосвязь данных классов представлена в виде UML-диаграммы на рисунке 12 [20].

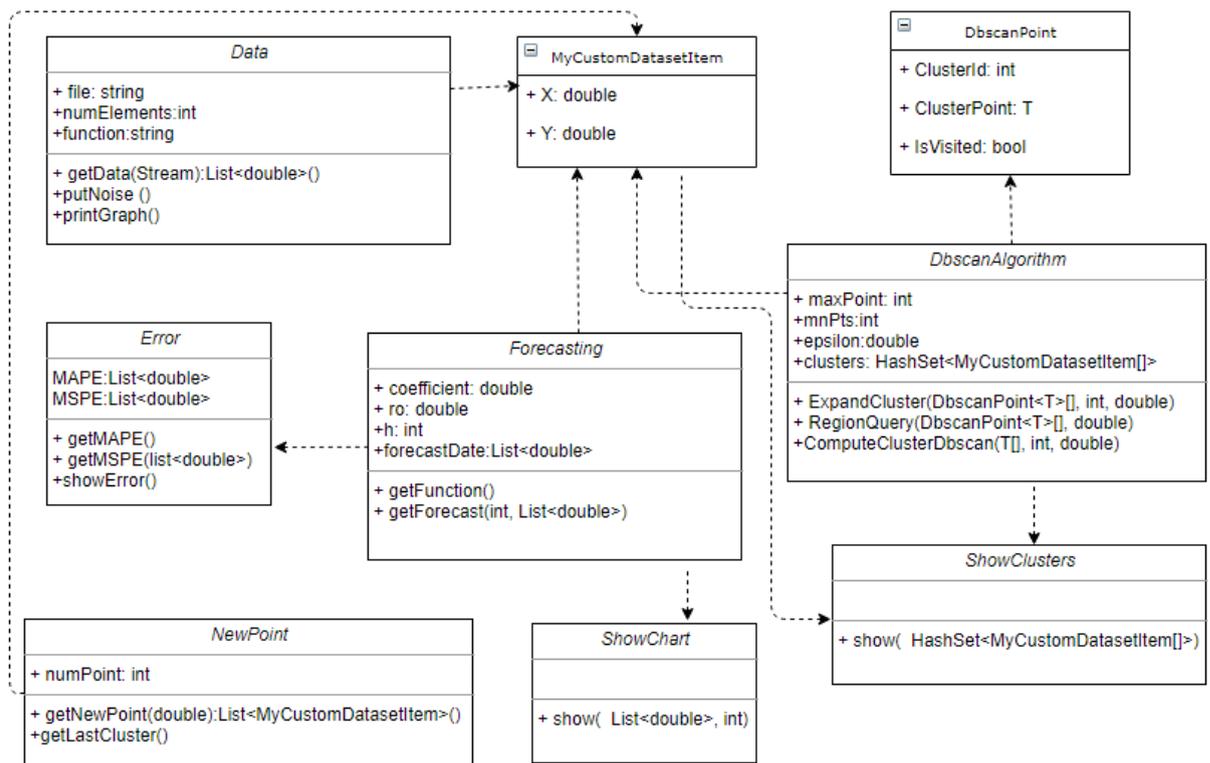


Рисунок 12 – UML-диаграмма классов

### 3.3. Функциональные требования к программному приложению

Реализованное программное приложение должно позволить пользователю построить краткосрочный прогноз по данным, представленным временным рядом.

Пользователь должен иметь возможность загрузить собственные данные путем выбора файла с нужным расширением с машины. Также, для получения навыков работы с приложением, апробации методов работы алгоритма, настройки параметров математического аппарата и т.д., необходима реализация возможности моделирования данных по некоторым функциям.

Пользователь должен иметь возможность настраивать некоторые параметры алгоритмов (кластеризация –  $\epsilon$ , кол-во элементов в наборе данных, прогнозирование -  $\rho$ ) вручную. При этом данные параметры должны иметь значения по умолчанию, отображаемые в соответствующих окнах.

Пользователь должен иметь возможность настраивать параметр  $h$  (количество элементов внутри звена сплайна) с помощью алгоритма

кластеризации. Результаты кластеризации должны отображаться пользователю в виде графика.

После настройки параметра  $h$  пользователю должно быть доступно получение прогноза по выбранным данным. Пользователю должно быть доступно добавление новых данных. В любой момент пользователь должен иметь возможность возобновить процесс получения прогноза с начальной порции данных. В любой момент пользователь должен иметь возможность выбора других данных для получения прогноза.

Пользователю должны быть доступны результаты построения прогноза и значений сглаживающего сплайна. Данная возможность должна быть реализована в виде графика.

При получении каждого нового значения прогноза пользователю должны быть доступна оценка качества прогноза в виде отображения ошибок прогноза (MAPE, MSPE).

Диаграмма вариантов использования представлена на рисунке 13.

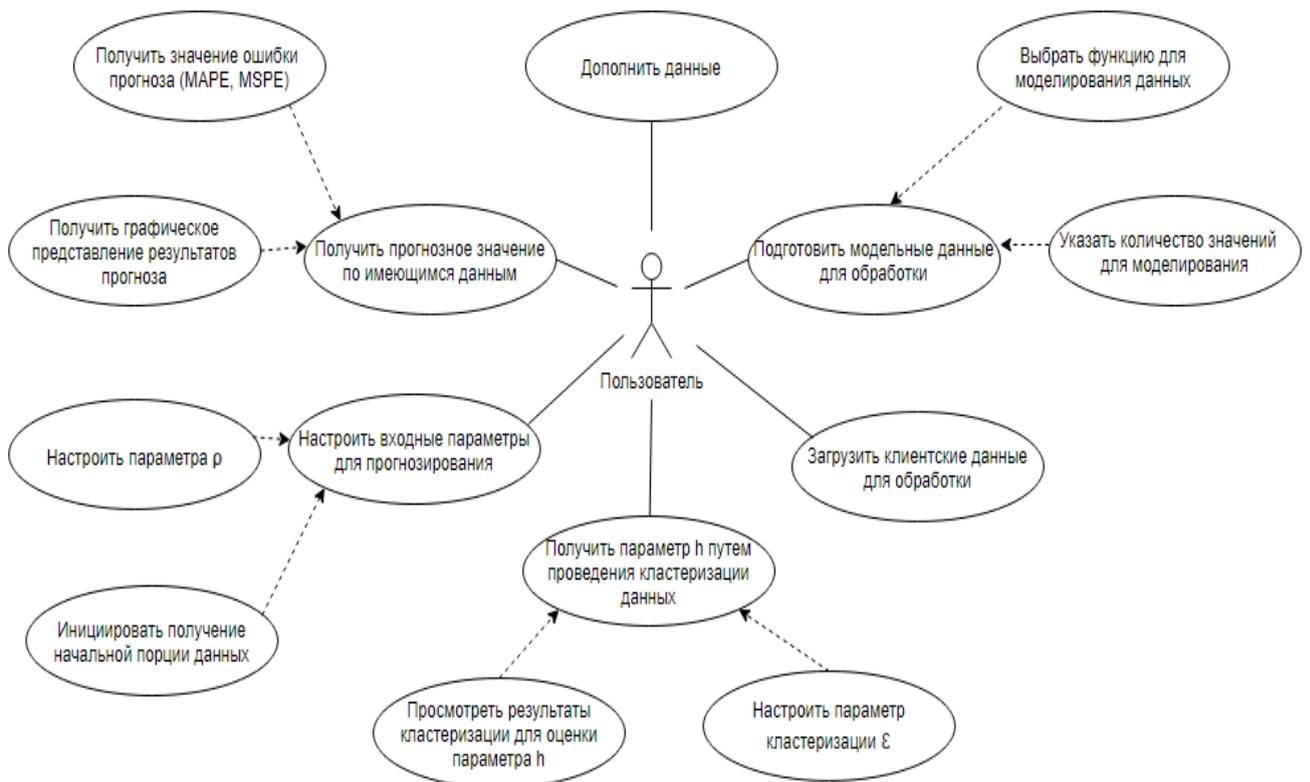


Рисунок 13 – Диаграмма вариантов использования для программного приложения

### 3.4. Интерфейс пользователя

Работу программного приложения тематически можно разделить на два этапа – настройка параметров с помощью кластерного анализа и построение прогноза. Для удобства работы пользователя данные функции были разнесены на два отдельных окна (две вкладки).

При запуске приложения пользователь будет направлен на вкладку настройки параметров (кластеризации). Здесь ему будет доступен выбор вида данных (пользовательские, модельные). Для работы с пользовательскими данными необходимо выбрать соответствующий файл. Для работы с модельными данными необходимо выбрать одну из предложенных функций в выпадающем списке и указать количество элементов. На этой же вкладке осуществляется настройка (ручная) параметров кластеризации. Результат кластеризации будет представлен пользователю в виде графика данных с различным цветовым отображением полученных кластеров (рис.14).

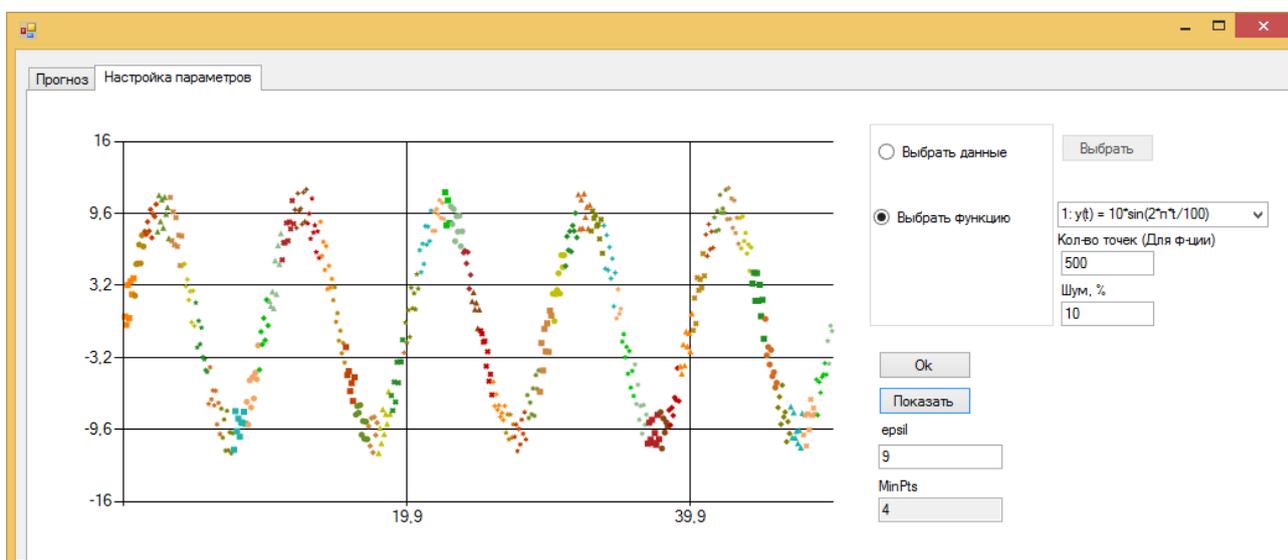


Рисунок 14 – Окно «Настройка параметров» с помощью кластеризации

После выбора данных пользователю будет доступен переход к вкладке «Прогноз». Здесь пользователь сможет инициировать получение начальной

выборки данных и настраивать (изменять) параметр  $\rho$ . Пользователю будет доступно пошаговое добавление данных для возможности просмотра динамики прогноза и оценки ошибки прогнозирования в каждой конкретной точке.

На этой же вкладке пользователю будет доступно графическое представление прогноза и история значений ошибок прогнозирования с первого полученного значения по текущее значение (рис.15).

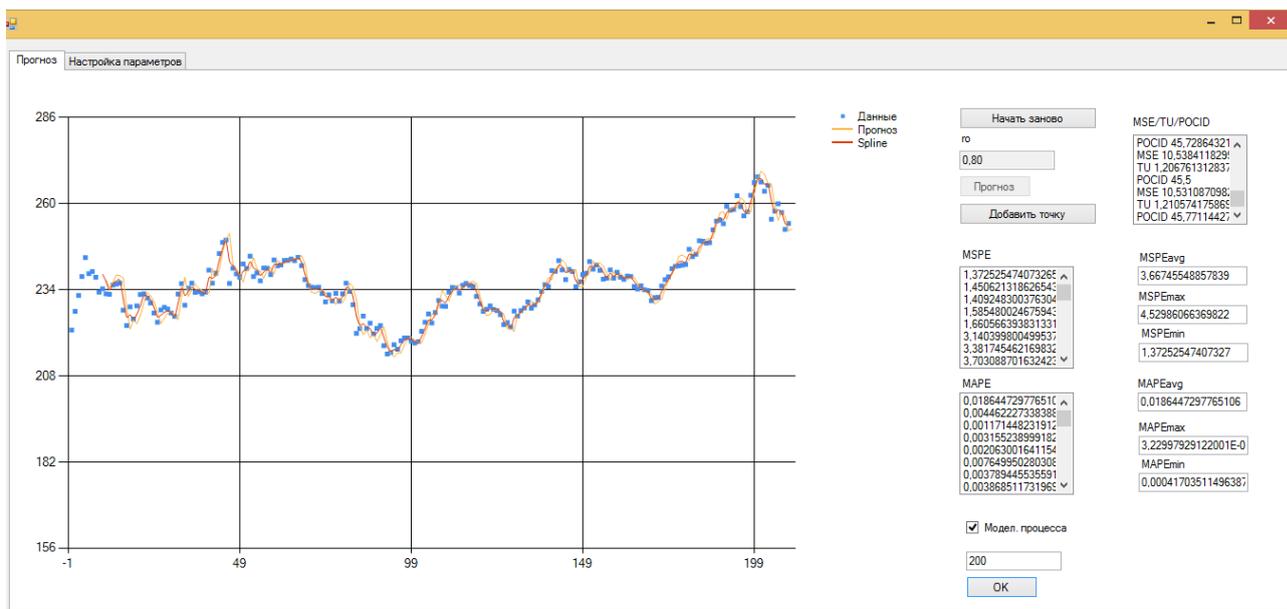


Рисунок 15 – Окно «Прогноз»

## 4. АНАЛИЗ РЕЗУЛЬТАТОВ ПРОГНОЗИРОВАНИЯ

### 4.1. Описание конечного алгоритма прогнозирования

Первый этап работы программы – обработка начальных данных (10 точек) и разбиение их на кластеры. Значения из последнего кластера являются входными данными для прогноза. Если в последнем кластере количество точек недостаточно (менее 4), необходимое количество точек добавляется из предыдущего кластера.

Определяются основные параметры прогноза:

- $h$  - количество точек в последнем кластере
- $m = h+1$  – точка, в которой получается значение-прогноз.

Далее (с учетом всех поступивших значений) пересчитываются коэффициенты  $A, B, C, a_0, a_1$  и т.д.

После настройки основных параметров аппарата построения прогноза были опробованы несколько способов получения прогноза:

- Функции  $F_1$  и  $F_2$  вычисляются с использованием поступивших значений (от 0 до  $h$ ), прогноз считается при  $m = h+1$ .
- Функции  $F_1$  и  $F_2$  вычисляются с использованием поступивших значений (от 0 до  $h$ ) и значением, смоделированным для точки  $m (h+1)$ . Прогноз считается при  $m = h+1$ .

### 4.2. Описание данных, участвующих в прогнозе

Для подготовки модельных данных используются функции, на которых были проведены тесты всех предыдущих алгоритмов (1, 2). На эти данные был наложен шум в заданное количество процентов от разницы максимального и минимального значения функции в определенном интервале (значение вводится пользователем в диапазоне от 0% до 100%).

Для представления реальных данных были использованы:

- Цены закрытия акций ПАО «Сбербанк» за период с 04.05.2019 по 04.05.2020 (рис.16);

- Цены закрытия акций круизной компании «Carnival Corporation» за период с 01.01.2020 по 05.05.2020 (рис.17);
- Цены закрытия акций ОАО «Мегафон» 2019 год (рис.18).

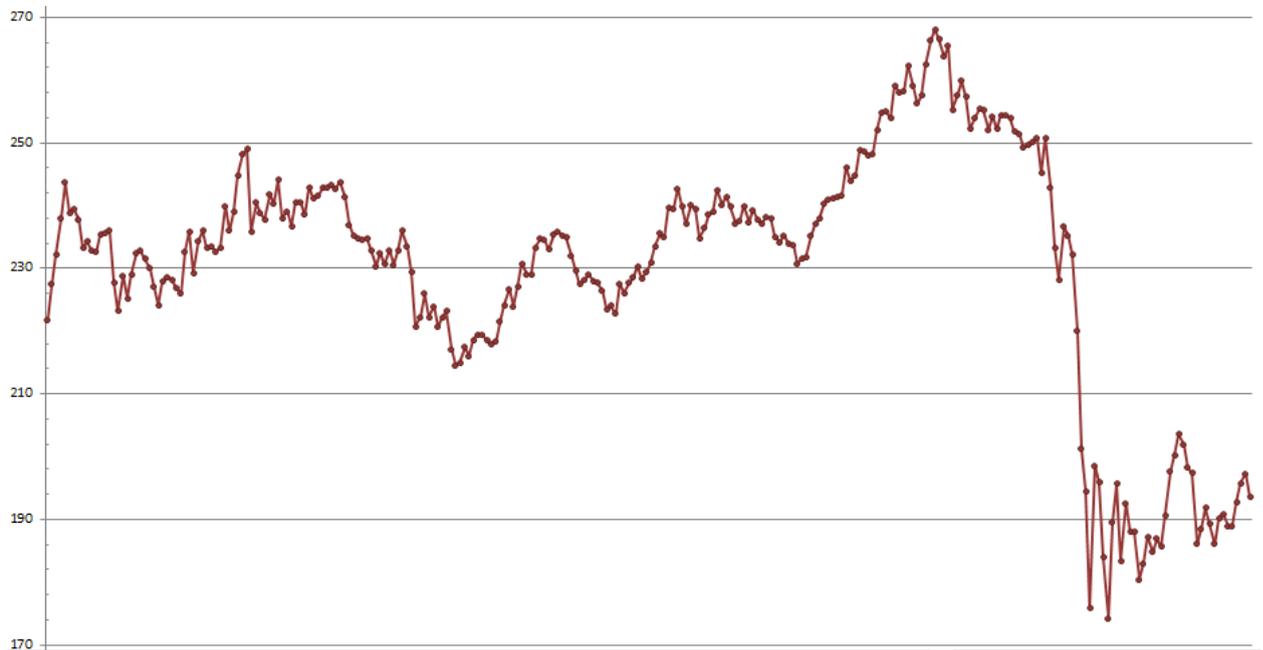


Рисунок 16 – Цены закрытия акций ПАО «Сбербанк»

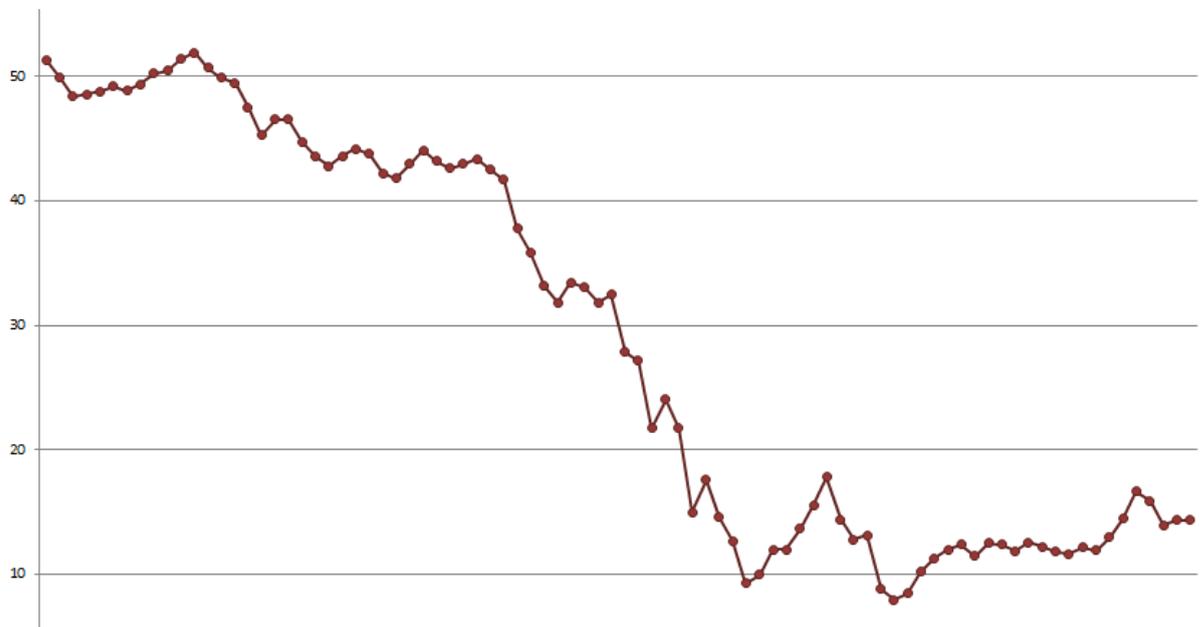


Рисунок 17 - Цены закрытия акций компании «Carnival Corporation»

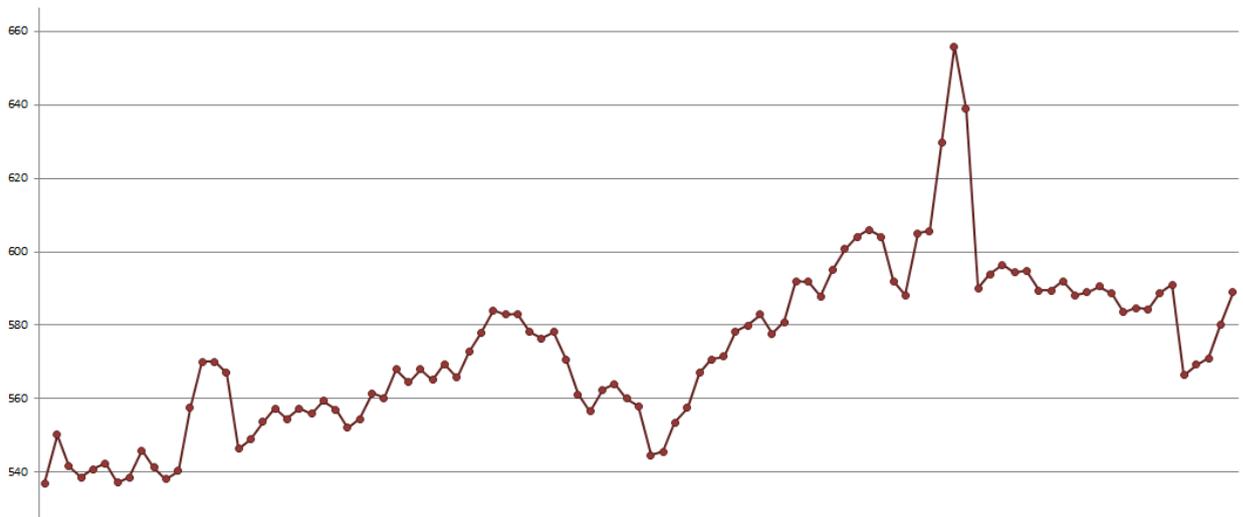


Рисунок 18 - Цены закрытия акций ОАО «Мегафон»

### 4.3. Результаты прогнозирования для модельных данных

Для оценки точности восстанавливаемой функции (и для оценки погрешности прогноза в дальнейшем) будет использоваться среднеквадратичная процентная ошибка. Данный способ оценивания помогает избежать несимметричности оценок. Для масштабируемости данной оценки каждое полученное значение будет отнесено к диапазону значений ряда. Формула для вычисления погрешности (MSPE):

$$MSPE = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (s_i - y_i)^2} \cdot \frac{100}{\max(y) - \min(y)} \quad (7)$$

Где  $n$  - количество измерений,  $s_i$  - полученное в результате прогнозирования значение в точке  $i$ ,  $y_i$  - действительное значение ряда в точке  $i$ .

На примере модельных данных, построенных по функциям №1-2 (5, 6), изучено влияние шума на результат прогноза. На подготовленные данные был наложен шум в 0%, 5%, 10% и 20% от разницы максимального и минимального значения функции на имеющемся интервале данных. Результаты представлены в таблицах 4 и 5.

Таблица 4 – Зависимость ошибки прогноза от значения шума для функции №1

$\sigma_{\xi}, \%$	0	5	10	20
$MSPE_{\min}$	0,469	0,354	2,692	1,073
$MSPE_{\max}$	3,065	2,233	3,970	6,169
$MSPE_{\text{avg}}$	2,343	1,984	3,777	5,570

Таблица 5 – Зависимость ошибки прогноза от значения шума для функции №2

$\sigma_{\xi}, \%$	0	5	10	20
$MSPE_{\min}$	0,369	0,231	0,946	3,008
$MSPE_{\max}$	3,580	4,190	6,617	8,274
$MSPE_{\text{avg}}$	1,738	2,381	4,354	6,090

Наборы данных, подготовленные для функций 1 и 2, содержат более 100 значений. Для настройки параметра  $h$  были подобраны и установлены параметры для DBScan ( $\text{eps} = 0.9$ ,  $\text{minPts} = 4$ ,  $\text{maxPts} = 10$ ). Результаты прогноза для функции №1 представлены на рисунках 19, 20 и в таблице 6.

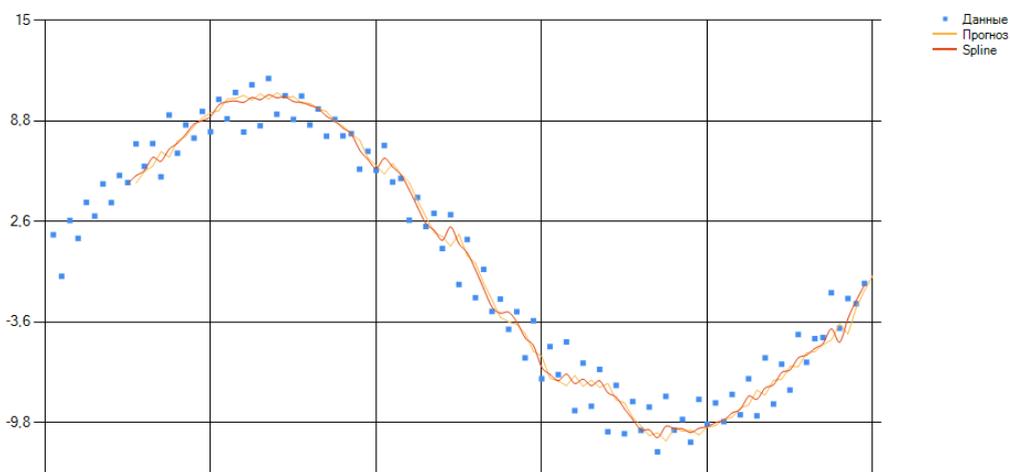


Рисунок 19 – Результаты построения прогноза для функции №1 с шумом в 10%

( $\rho = 0.2$ )

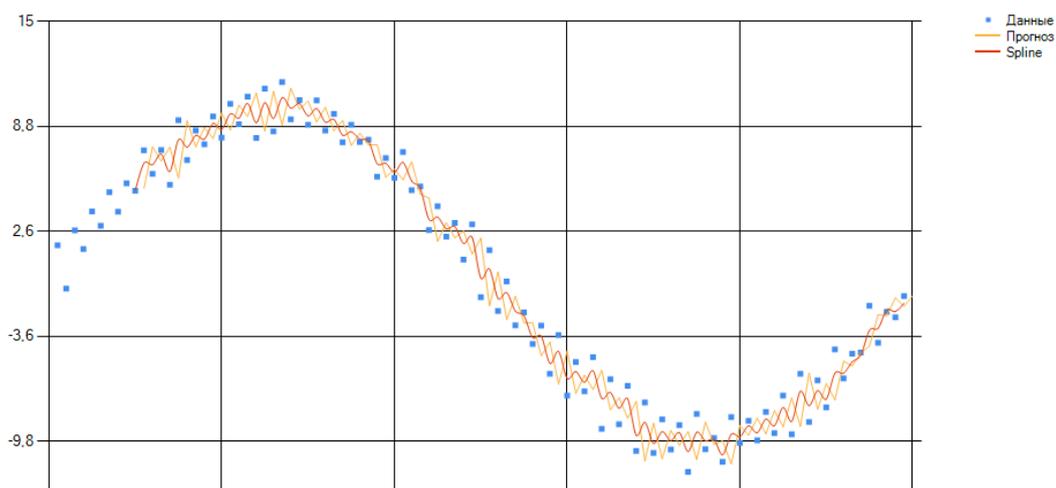


Рисунок 20 – Результаты построения прогноза для функции №1 с шумом в 10%  
( $\rho = 0.8$ )

Таблица 6 – Результаты прогноза для функции №1

$\rho$	0,2	0,4	0,6	0,8
$MSPE_{\min}$	1,259	1,117	1,112	1,025
$MSPE_{\max}$	2,464	2,264	2,749	3,720
$MSPE_{\text{avg}}$	2,143	1,842	2,354	3,310

Анализ полученных результатов показал, что наилучшие результаты прогноза были выявлены при значении параметра  $\rho = 0,4$ . Максимальная среднеквадратическая ошибка при неоптимальных параметрах не превышает 4%. Данные показатели превосходят по качеству показатели, полученные при настройке параметров с помощью генетического алгоритма. При использовании оптимальных параметров при работе с реализованной моделью прогнозирования максимальное значение среднеквадратической ошибки (2,26) оказывается меньшим, нежели минимальное значение среднеквадратической ошибки (3,22) при настройке параметров генетическим алгоритмом.

Результаты прогноза для функции №2 представлены на рисунках 21, 22 и в таблице 7.

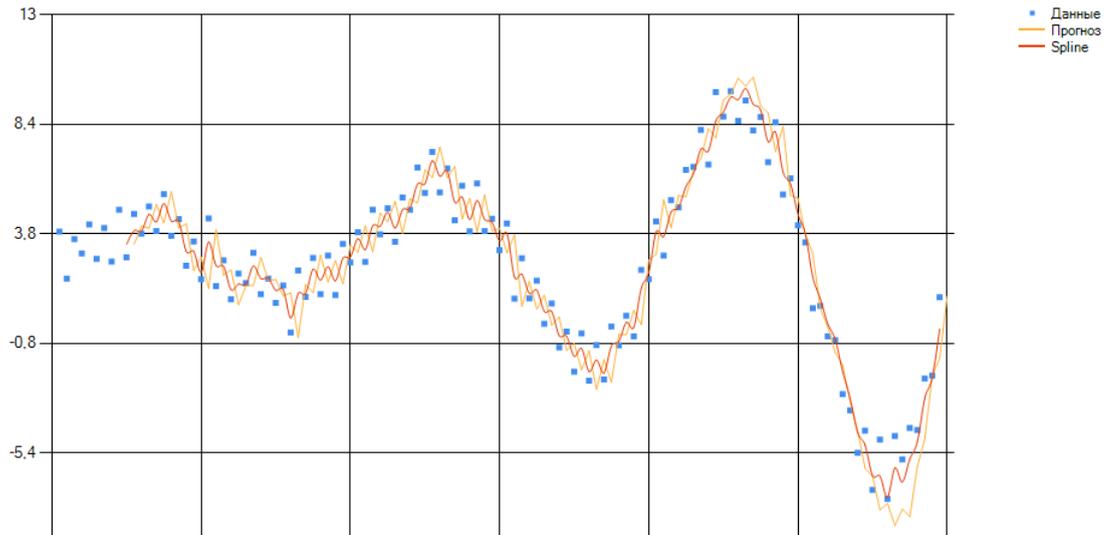


Рисунок 21 – Результаты построения прогноза для функции №2 с шумом в 10% ( $\rho = 0.8$ )

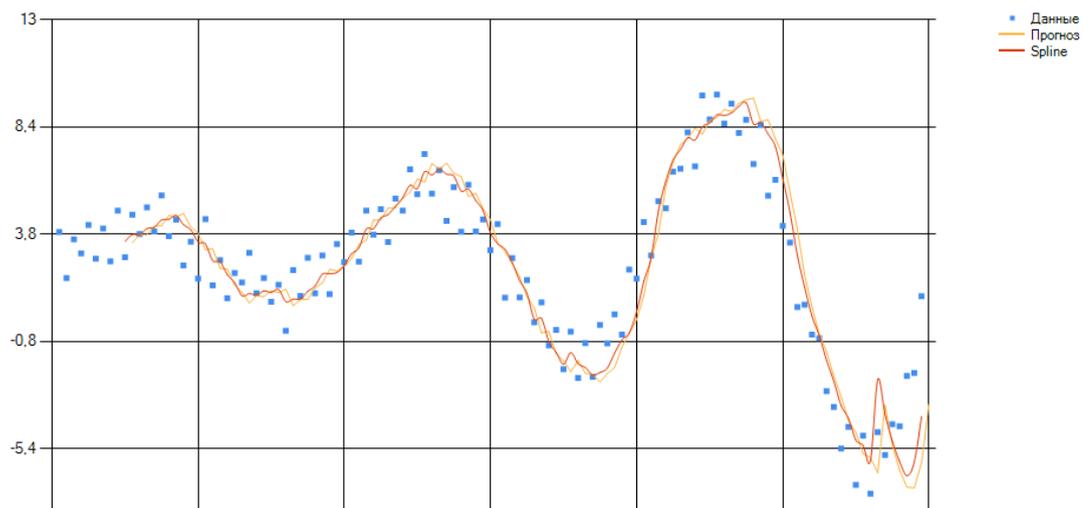


Рисунок 22 – Результаты построения прогноза для функции №2 с шумом в 10% ( $\rho = 0.2$ )

Таблица 7 – Результаты прогноза для функции №2

$\rho$	0,2	0,4	0,6	0,8
$MSPE_{\min}$	1,465	1,657	1,696	1,503
$MSPE_{\max}$	7,862	5,893	5,352	5,472
$MSPE_{\text{avg}}$	3,624	3,305	3,616	4,249

Анализ полученных результатов показал, что наилучшие результаты прогноза были выявлены при значении параметра  $\rho = 0,4$ . При этом значения ошибки прогноза для различных параметров  $\rho$  отличаются незначительно (не более 1%). Максимальная среднеквадратическая ошибка при неоптимальных параметрах не превышает 8%. Данные показатели превосходят по качеству показатели, полученные при настройке параметров с помощью генетического алгоритма. При использовании оптимальных параметров при работе с реализованной моделью прогнозирования максимальное значение среднеквадратической ошибки (5,35) оказывается близким к минимальному значению среднеквадратической ошибки (5,21) при настройке параметров генетическим алгоритмом. Минимальное значение среднеквадратической ошибки меньше данного значения.

#### **4.4. Результаты прогнозирования для реальных данных**

При получении каждого нового значения прогноза будет вычислена среднеквадратичная погрешность прогнозирования (MSPE). Помимо этого, для каждой точки будет считаться средняя абсолютная ошибка прогноза (MAPE):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - s_i|}{y_i} \quad (8)$$

Данная оценка зачастую используется для оценки точности прогноза. Оценка MAPE показывает, насколько велики ошибки в сравнении со значениями ряда. Данная оценка применяется для временных рядов, фактические значения которых значительно больше 1. Например, оценки ошибки прогнозирования энергопотребления зачастую приводятся как значения MAPE.

##### **4.4.1. Результаты прогнозирования для данных ПАО «Сбербанк»**

Набор данных, подготовленный по ценам акций закрытия ПАО «Сбербанк», содержит более 250 значений. Для настройки параметра  $h$  были

подобраны и установлены параметры для DBScan ( $\text{eps} = 0.9$ ,  $\text{minPts} = 4$ ,  $\text{maxPts} = 10$ ). Результат кластеризации по данным параметрам представлен на рисунке 23. Результаты прогноза представлены на рисунке 24 и в таблице 8.

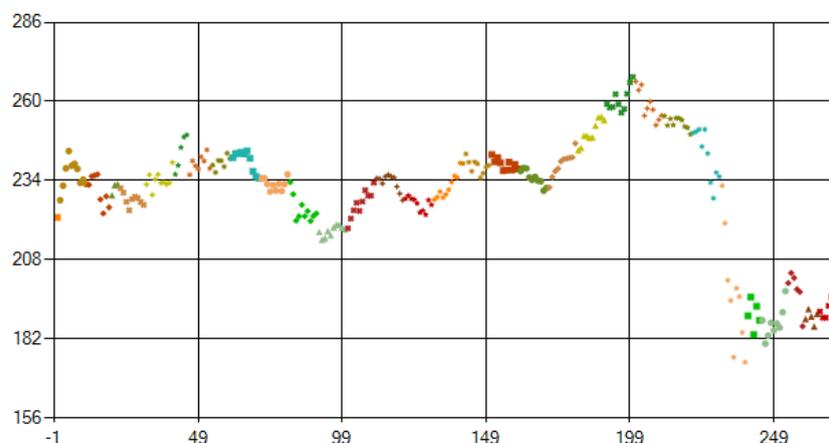


Рисунок 23 – Результат кластерного анализа данных ПАО «Сбербанк»

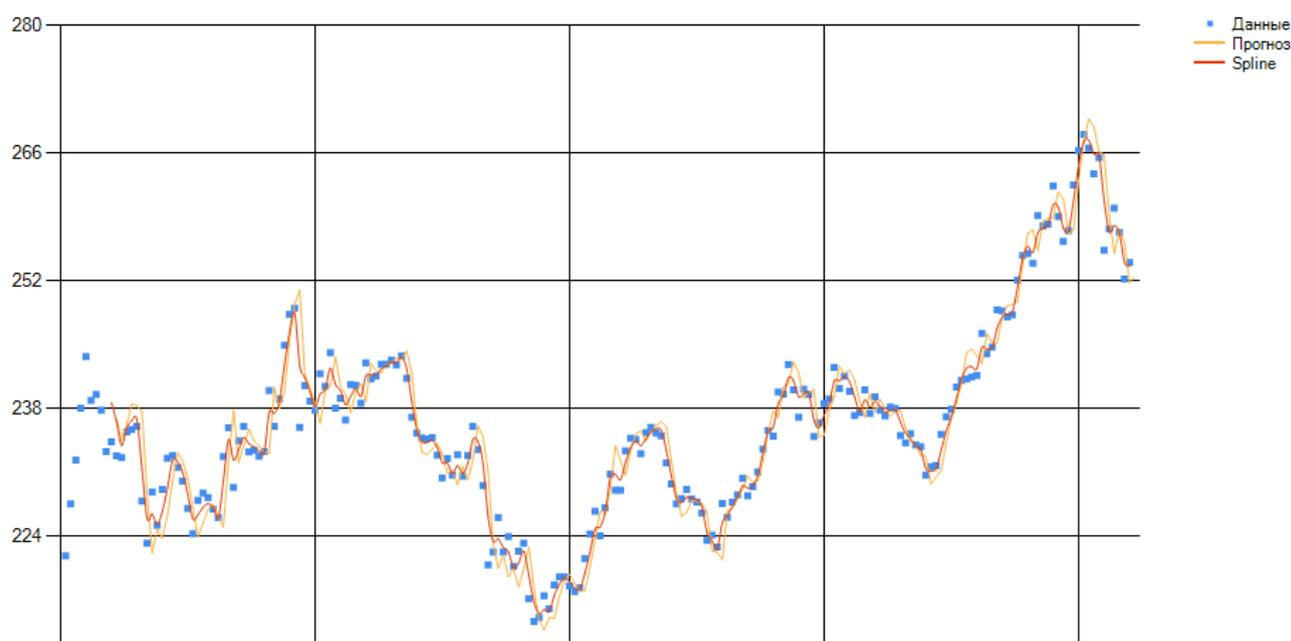


Рисунок 24 – Результат построения прогноза по данным ПАО «Сбербанк»

Таблица 8 – Результаты прогноза по данным ПАО «Сбербанк»

$\rho$	0,2	0,4	0,6	0,8
$\text{MSPE}_{\min}$	2,677	1,372	2,634	1,372
$\text{MSPE}_{\max}$	5,672	4,529	4,867	4,529
$\text{MSPE}_{\text{avg}}$	4,139	3,667	4,242	3,667
$\text{MAPE}_{\min}$	0,0001	0,0003	0,0002	0,0003

MAPE <sub>max</sub>	3,411	1,748	3,356	1,748
MAPE <sub>avg</sub>	0,055	0,039	0,057	0,039

Анализ полученных результатов показал, что наилучшие результаты прогноза были выявлены при значениях параметра  $\rho$  равных 0,6 и 0,8. Максимальная среднеквадратическая ошибка при неоптимальных параметрах не превышает 6%, средняя абсолютная ошибка прогноза не превышает 3,5%. Данные показатели превосходят по качеству показатели, полученные при настройке параметров с помощью генетического алгоритма.

#### 4.4.2. Результаты прогнозирования для данных круизной компании «Carnival Corporation»

Набор данных, подготовленный по ценам акций закрытия международной круизной компании «Carnival Corporation», содержит 100 значений. Для настройки параметра  $h$  были подобраны и установлены параметры для DBScan ( $\text{eps} = 0.4$ ,  $\text{minPts} = 4$ ,  $\text{maxPts} = 10$ ). Результат кластеризации по данным параметрам представлен на рисунке 25. Результаты прогноза представлены на рисунке 26 и в таблице 9.

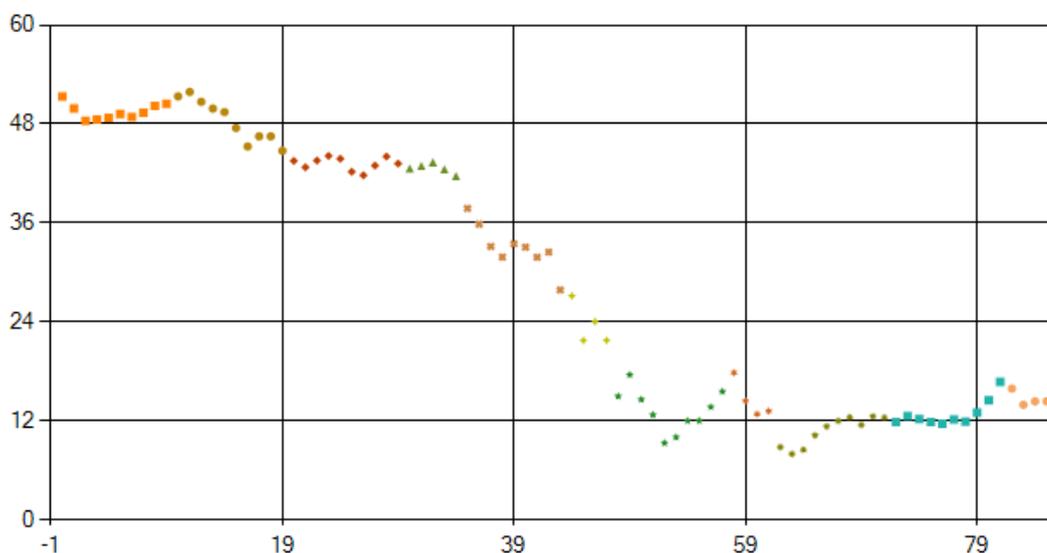


Рисунок 25 – Результат кластерного анализа данных компании «Carnival Corporation»

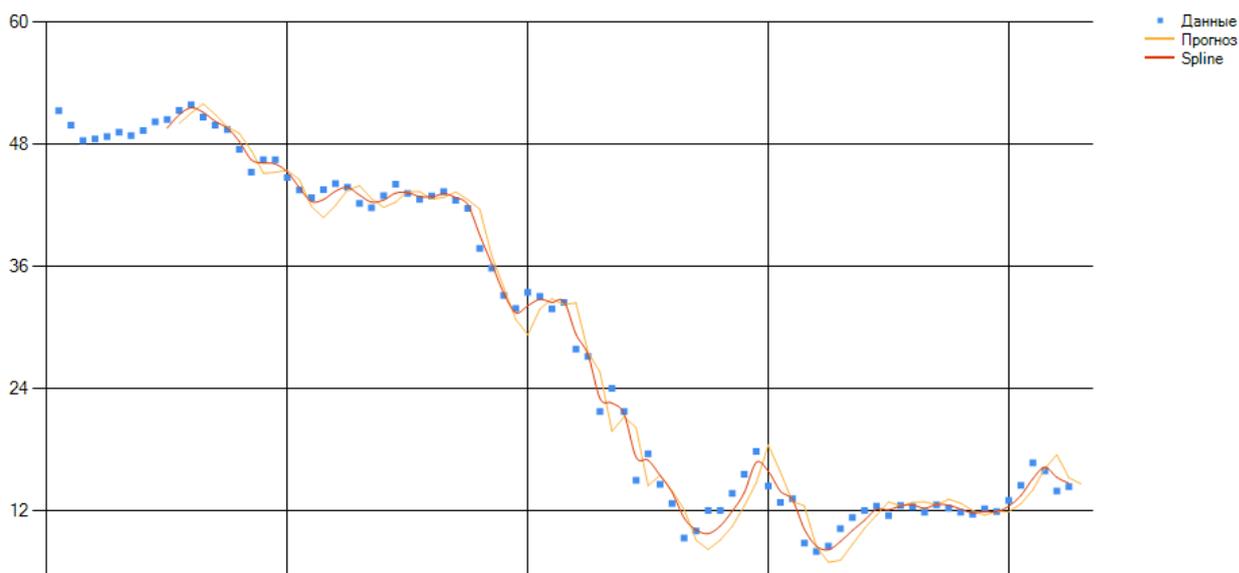


Рисунок 26 – Результат построения прогноза по данным компании «Carnival Corporation»

Таблица 9 – Результаты прогноза по данным компании «Carnival Corporation»

$\rho$	0,2	0,4	0,6	0,8
$MSPE_{\min}$	1,288	1,501	1,413	0,928
$MSPE_{\max}$	6,233	5,393	5,014	4,777
$MSPE_{\text{avg}}$	4,740	4,053	3,757	3,520
$MAPE_{\min}$	0,0013	0,006	0,002	0,0003
$MAPE_{\max}$	3,485	4,061	3,822	2,512
$MAPE_{\text{avg}}$	0,359	0,343	0,329	0,296

Анализ полученных результатов показал, что наилучшие результаты прогноза были выявлены при значении параметра  $\rho$  равном 0,8. При этом значения ошибки прогноза для различных параметров  $\rho$  отличаются незначительно (не более 1,3%). Максимальная среднеквадратическая ошибка при неоптимальных параметрах не превышает 6,5%, средняя абсолютная ошибка прогноза в районе 4%.

#### 4.4.3. Результаты прогнозирования для данных ОАО «Мегафон»

Набор данных, подготовленный по ценам акций закрытия ОАО «Мегафон», содержит более 100 значений. Для настройки параметра  $h$  были

подобраны и установлены параметры для DBScan ( $\text{eps} = 0.7$ ,  $\text{minPts} = 4$ ,  $\text{maxPts} = 10$ ). Результат кластеризации по данным параметрам представлен на рисунке 27. Результаты прогноза представлены на рисунке 28 и в таблице 10.

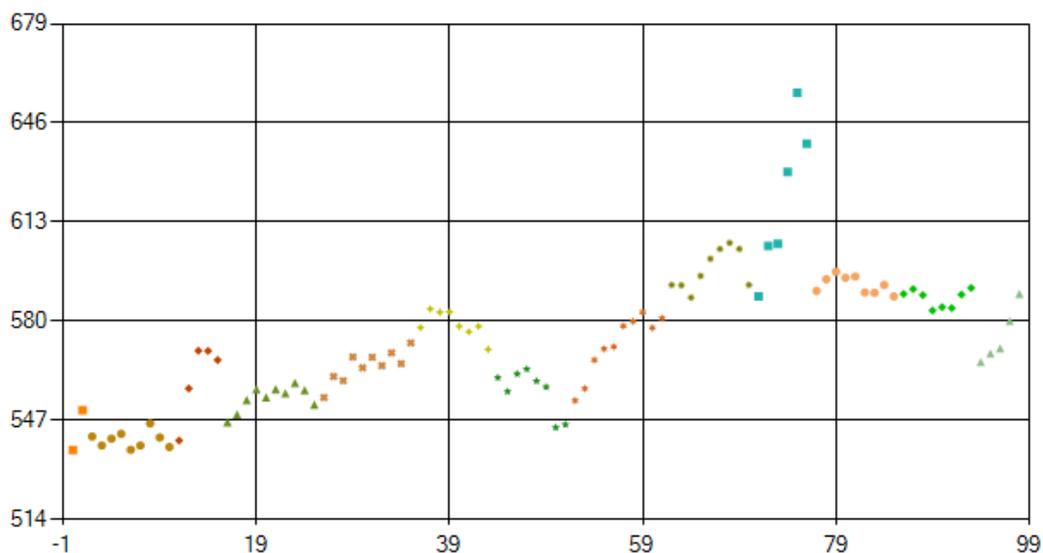


Рисунок 27 – Результат кластерного анализа данных ОАО «Мегафон»

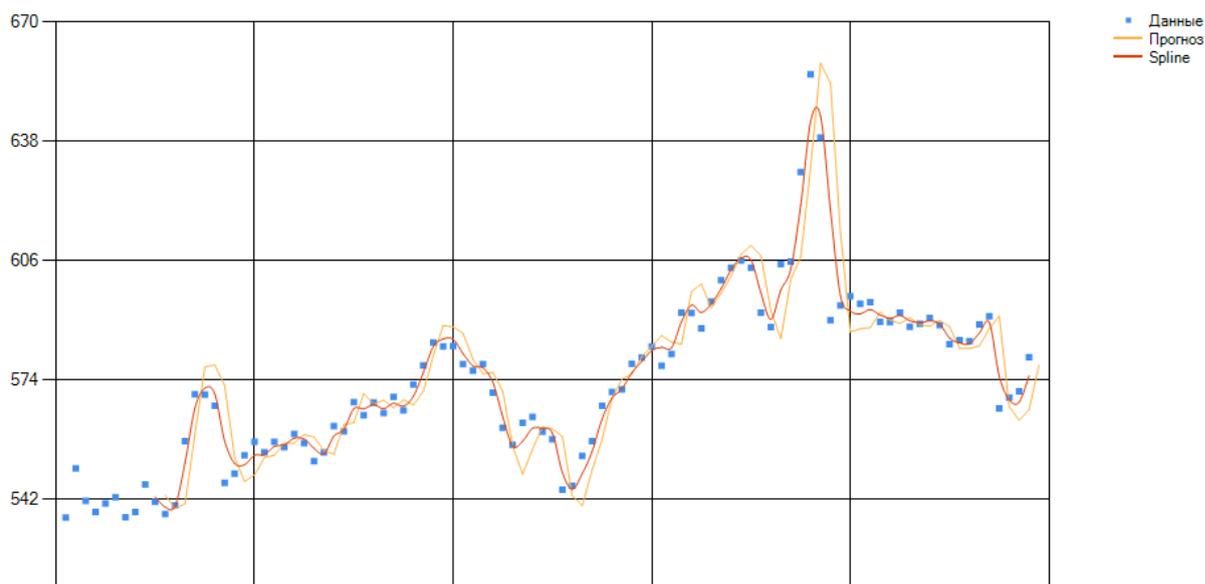


Рисунок 28 – Результат построения прогноза по данным ОАО «Мегафон»

Таблица 10 – Результаты прогноза по данным ОАО «Мегафон»

P	0,2	0,4	0,6	0,8
$\text{MSPE}_{\text{max}}$	11,612	10,454	9,856	9,458
$\text{MSPE}_{\text{min}}$	0,972	1,227	1,306	1,279
$\text{MSPE}_{\text{avg}}$	9,207	7,816	7,118	6,693

MAPE <sub>min</sub>	0,0006	0,0005	0,0002	0,0002
MAPE <sub>max</sub>	1,097	1,021	1,000	1,004
MAPE <sub>avg</sub>	0,091	0,077	0,071	0,069

Анализ полученных результатов показал, что наилучшие результаты прогноза были выявлены при значении параметра  $\rho = 0,8$ . Максимальная среднеквадратическая ошибка при неоптимальных параметрах не превышает 12%, средняя абсолютная ошибка прогноза не превышает 2%. Данные показатели превосходят по качеству показатели, полученные при настройке параметров с помощью генетического алгоритма.

#### **4.5. Анализ результатов прогнозирования по данным репозитория ICMC-USP Time Series Prediction**

Оценивание результатов работы метода производилось на открытых данных, используемых в статье [12] для сравнения известных методов прогнозирования. В данной статье используется набор данных из открытого онлайн-репозитория, в котором находится набор данных из 40 ВР: 17 детерминированных ВР, 15 стохастических ВР и 8 хаотических ВР. Временные ряды содержат от 450 до 3000 значений, смоделированные с помощью определенных математических моделей. Данные были опубликованы научными сотрудниками университета Сан-Паулу (USP), Бразилия.

##### **4.5.1. Показатели**

Для анализа результатов работы полученного метода прогнозирования необходимо не только оценивать ошибки прогноза, но и сравнивать результаты работы метода с результатами других методов (алгоритмов) прогнозирования. Сравнение алгоритмов может быть сложной задачей, когда доступны разные показатели производительности. Метод, который содержит различные релевантные переменные, является ключевым элементом для решения этой проблемы.

Помимо ошибки прогнозирования, выраженной в разнице фактических и прогнозируемых значений, существуют другие методы оценки прогноза. Так,

например, коэффициент Тейла (TU) показывает производительность исследуемого алгоритма. Данный коэффициент принимает значения в интервале от 0 (наилучшее значение) до  $+\infty$  (наихудшее значение). Рассчитанные значения коэффициента можно интерпретировать следующим образом: если  $TU > 1$ , производительность алгоритма ниже, чем наивной модели; если производительность алгоритма такая же, как у наивной модели; если  $TU < 1$ , производительность алгоритма выше, чем наивной модели; и если  $TU \leq 0,55$ , алгоритм, представляющий интерес, может выполнять будущие прогнозы.

$$TU = \frac{\sum_{i=1}^n (y_i - s_i)^2}{\sum_{i=1}^n (y_i - y_{i-1})^2} \quad (9)$$

Другой рассматриваемый индекс производительности – ROCID:

$$ROCID = \frac{\sum_{i=1}^n D_i}{n} \cdot 100 \quad (10)$$

Термин  $D_i$  хранит значение 1, если выполняется условие  $(y_i - y_{i-1})(s_i - s_{i-1}) > 0$  и 0 в противном случае. Идея этого индекса состоит в том, чтобы оценить точность изменений направления прогнозируемых данных, т. е. будет ли будущее значение увеличиваться или уменьшаться по сравнению с текущим значением. Индекс ROCID может быть использован в качестве дополнительного способа анализа ошибок прогнозирования.

Для сравнения работы методов также используется мера MSE. Данный показатель работает с разницей (ошибкой) между фактическим наблюдаемым значением и прогнозируемым значением. Данный показатель принимает значения в интервале от 0 (наилучшее значение) до  $+\infty$  (наихудшее значение). MSE – среднеквадратическая ошибка прогноза применяется в ситуациях, когда нам надо подчеркнуть большие ошибки и выбрать модель, которая дает меньше

больших ошибок прогноза. Грубые ошибки становятся заметнее за счет того, что ошибку прогноза мы возводим в квадрат.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - s_i)^2 \quad (11)$$

Использование векторного показателя, основанного на всех вышеперечисленных индексах, поможет сравнить результаты прогноза полученного метода с результатами известных методов (ARIMA, LSTM и т.д.) и сделать выводы о его пригодности. Для получения векторного показателя индексы используются с некоторыми коэффициентами, определяющими значимость конкретного индекса в общей оценке прогноза. При этом в отличие от показателей ошибок, которые дают значения, которые должны быть минимизированы, ROCID должен быть максимизирован. Для каждого параметра рассчитывается среднее значение по всем временным рядам в категории.

#### **4.5.2. Синтетические данные**

В статье [12] приводятся результаты экспериментов на коллекции, которая охватывает 40 синтетических временных рядов, из которых 17 являются детерминированными, 15 стохастическими (случайными) и 8 хаотическими. Далее будут детализированы результаты прогнозирования для отдельных категорий ВР.

##### **4.5.2.1. Детерминированные временные ряды**

Результаты сравнения индекса TU представлены на рисунке 29. Результаты показывают для каждого метода, сколько наборов данных (в процентах) представили значения TU в одном из трех возможных диапазонов. Можно отметить, что SARIMA показывает наибольшую производительность - в 14 из 17 наборов данных индекс TU не превышает оптимального значения ( $TU \leq 0,55$ ). Слабый результат производительности показывают методы HES и SES – на всех наборах данных индекс TU превысил критическое значение ( $TU > 1$ ).

Полученный метод (P-spline) показывает средние результаты – доля рядов, у которых индекс производительности TU не превышает критическое значение, превышает 30% (5-ое место по производительности среди остальных методов).

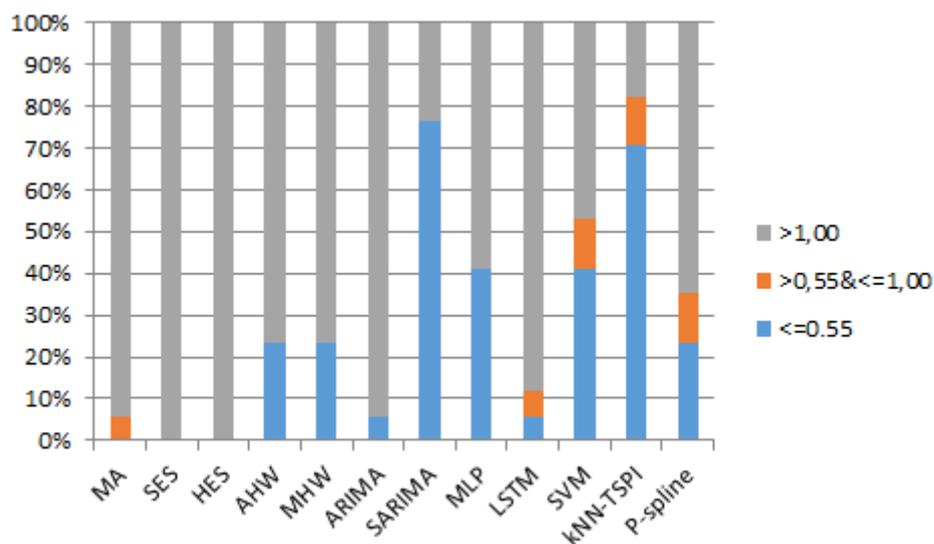


Рисунок 29 – Распределение результатов TU по диапазонам

Для детерминированных временных рядов также были рассчитаны ошибка MSE и индекс ROCID. С их учетом для каждого метода был рассчитан векторный показатель (рис.30). Наименьшее значение показателя указывает на наилучший результат работы алгоритма (ошибки, производительность и т.д.). На данной категории временных рядов полученный метод показывает низкие результаты (10-ое место). Большой вклад в конечное значение векторного показателя вносит ошибка MSE - алгоритм показывает неудовлетворительный результат на данных с выбросами в большом диапазоне.

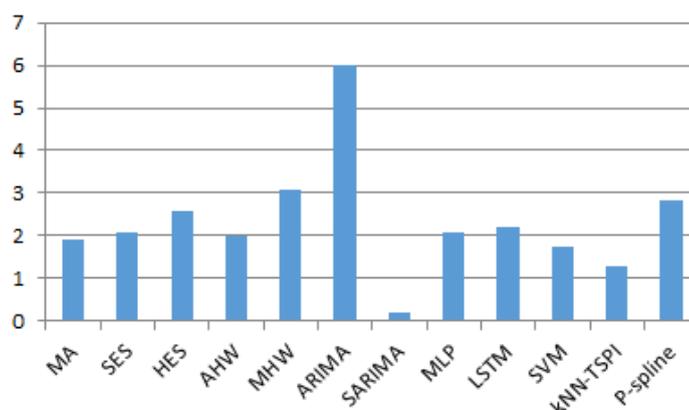


Рисунок 30 – Значения векторного показателя для детерминированных ВР

#### 4.5.2.2. Случайные временные ряды

Результаты сравнения индекса TU для случайных временных рядов (рис.31) показывают, что несколько методов (SARIMA, SVM, ARIMA) показывают наибольшую производительность. Худший результат производительности показывает метод HES – на всех наборах данных индекс TU превысил критическое значение ( $TU > 1$ ). Полученный метод (P-spline) показывает средние результаты – доля рядов, у которых индекс производительности TU не превышает критическое значение, превышает 40%, но при этом нет оптимальных значений индекса (6-ое место по производительности среди остальных методов).

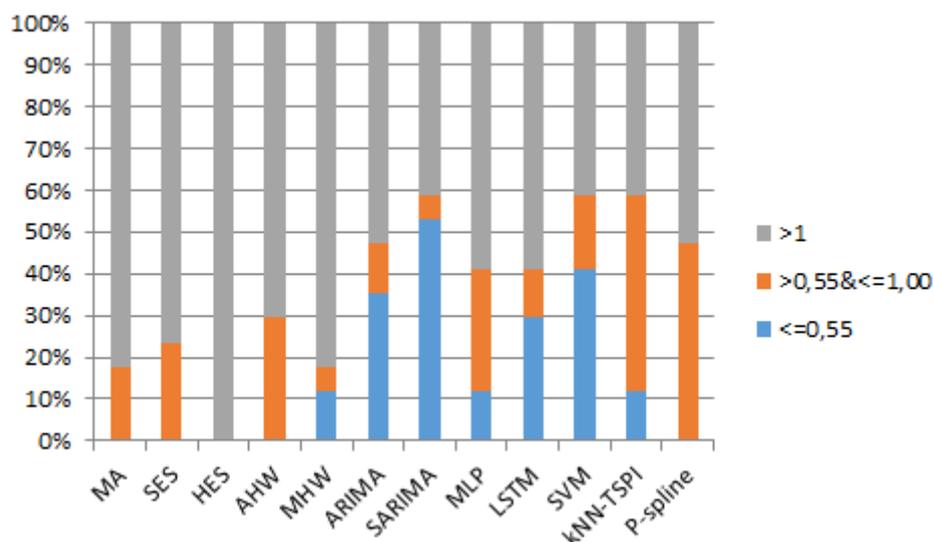


Рисунок 31 – Распределение результатов TU по диапазонам

Для детерминированных временных рядов также были рассчитаны ошибка MSE и индекс ROCID. С их учетом для каждого метода был рассчитан векторный показатель (рис.32). Для случайных временных рядов полученный метод показывает высокие результаты (разделяет 2-ое место с MA).

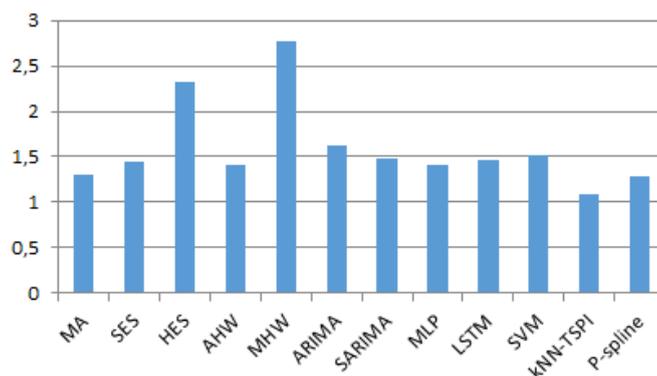


Рисунок 32 – Значения векторного показателя для случайных ВР

### 4.5.2.3. Хаотические временные ряды

По результатам сравнения индекса TU для хаотических временных (рис.33) невозможно выделить метода с высокой производительностью - доля рядов с низким индексом производительности больше 75%. Худший результат производительности показывают методы HES и AHW – на всех наборах данных индекс TU превысил критическое значение ( $TU > 1$ ). Полученный метод (P-spline) показывает низкие результаты – доля рядов, у которых индекс производительности TU не превышает критическое значение, примерно 13%, при этом нет оптимальных значений индекса (9-ое место по производительности среди остальных методов).

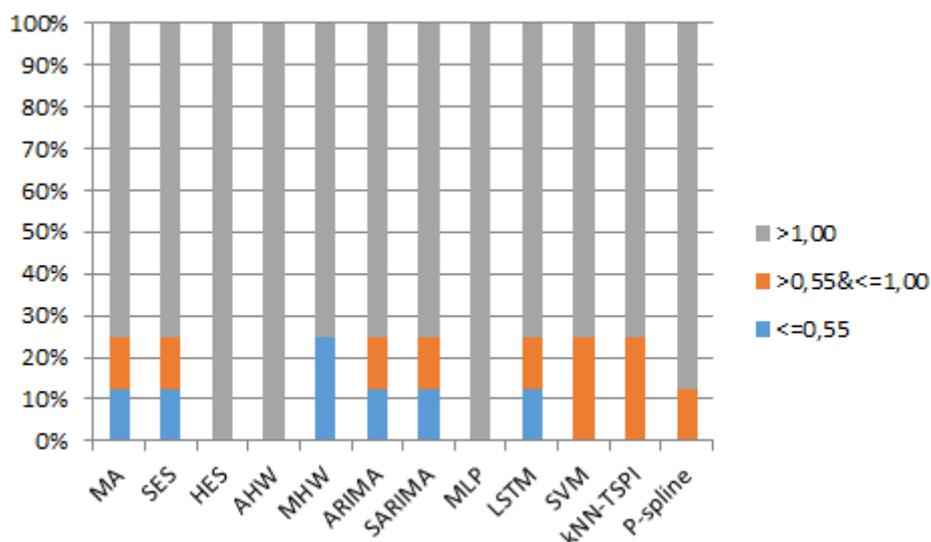


Рисунок 33 – Распределение результатов TU по диапазонам

Для детерминированных временных рядов также были рассчитаны ошибка MSE и индекс ROCID. С их учетом для каждого метода был рассчитан векторный показатель (рис.34). Полученный метод близок по значению векторного показателя к нескольким методам (MHW, SARIMA, AHW).

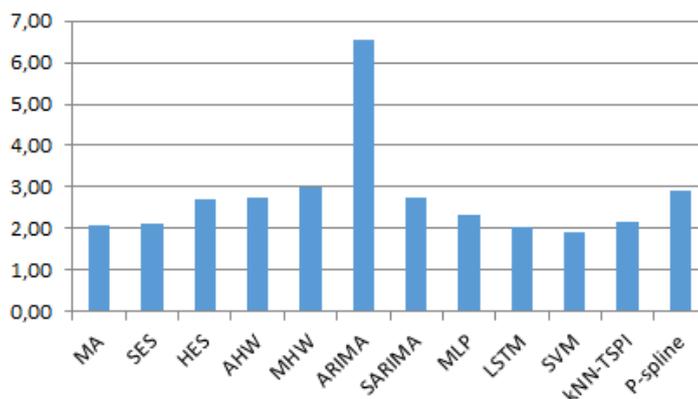


Рисунок 34 – Значения векторного показателя для хаотических ВР

### 4.5.3. Общие результаты сравнения методов

На рисунках 35-37 представлены результаты сравнения ошибки MSE и индексов производительности TU и ROCID для всех методов прогнозирования по всем категориям ВР.

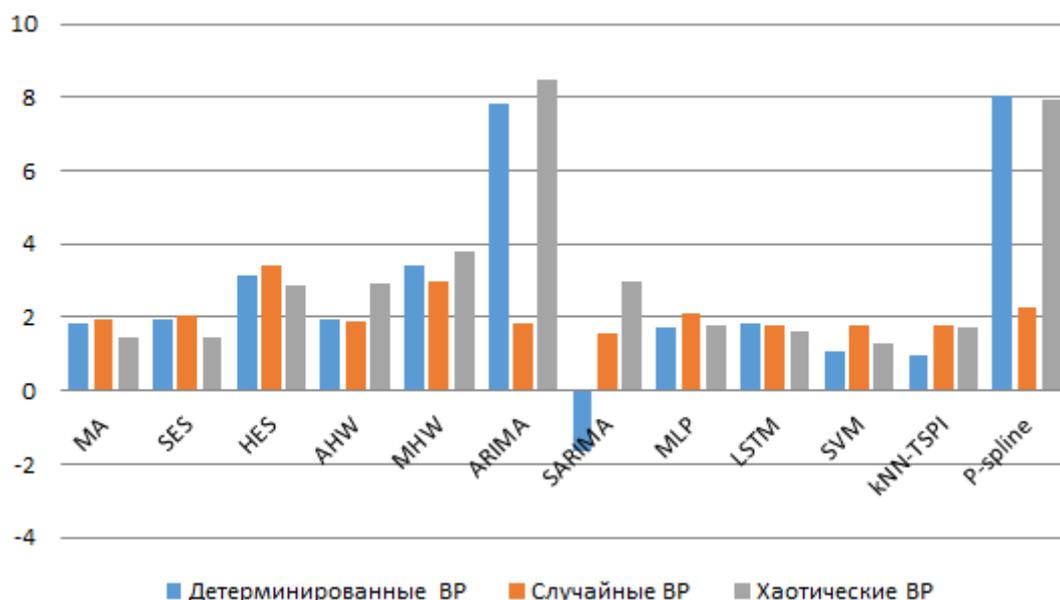


Рисунок 35 – Среднее значение MSE для всех категорий ВР

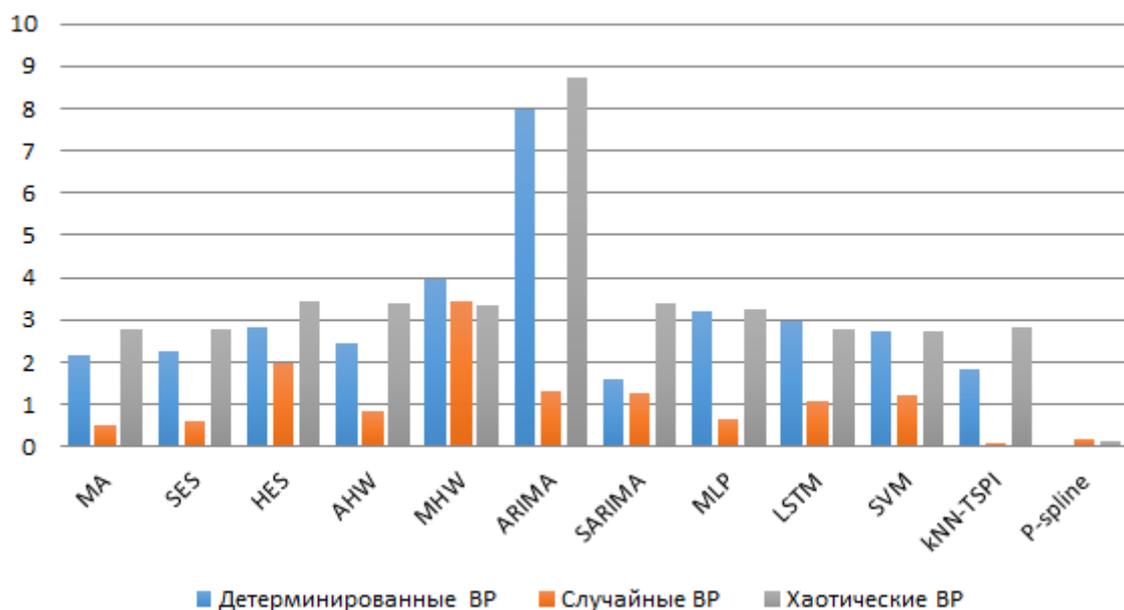


Рисунок 36 - Среднее значение TU для всех категорий BP

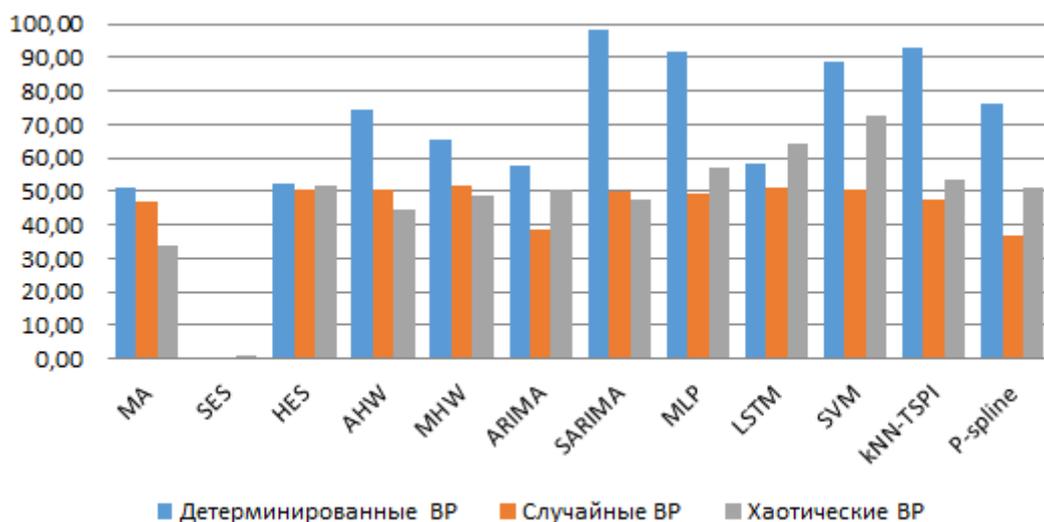


Рисунок 37 - Среднее значение POCID для всех категорий BP

Векторный показатель для каждого метода рассчитывается следующим образом:

$$V_i = 0,3 \cdot \log MSE_i + 0,4 \cdot \log TU_i + 0,3 \cdot \log(100 - POCID_i) \quad (12)$$

Векторный показатель принимает значения от 0 до  $+\infty$ .

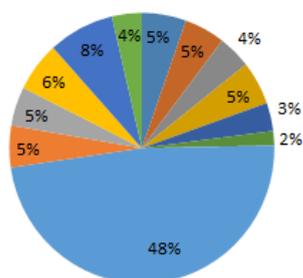
Обратные значения вычисленных векторных показателей, представленные в процентах (от 0 до 100%), показаны в таблице 11 и на рисунке 38. Наибольшее значение в группе указывает на наилучший результат работы алгоритма.

Так, например, для хаотических ВР наилучший показатель в группе имеет алгоритм SVM (11,04), худший результат показывает алгоритм ARIMA (3,19).

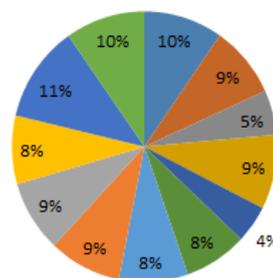
Таблица 11 – Сравнение векторных показателей

	Детерминированные ВР	Случайные ВР	Хаотические ВР
MA	5,396299	9,618233	9,998339598
SES	4,959128	8,618956	9,776650751
HES	4,026845	5,358725	7,685595097
AHW	5,217371	8,902322	7,594213384
MHW	3,366423	4,500898	7,025449682
ARIMA	1,72115	7,681386	3,197789273
SARIMA	48,01509	8,424353	7,58958664
MLP	4,984959	8,849327	8,992609984
LSTM	4,67738	8,568924	10,22603287
SVM	5,960279	8,203351	11,0414577
kNN-TSPI	8,023396	11,55758	9,735648058
P-spline	3,651678	9,715946	7,136626954

Детерминированные ВР



Случайные ВР



Хаотические ВР

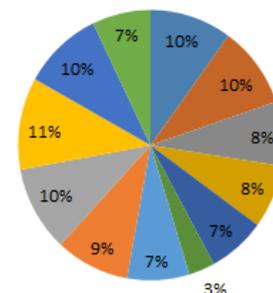


Рисунок 38 – Векторные показатели алгоритмов по всем категориям ВР

Сравнение полученных результатов показало, что полученный метод уступает большинству методов (ARIMA имеет схожий результат) по среднеквадратической ошибке прогноза MSE. Значения индекса ROCID, рассчитанные для полученного метода, схожи со значениями большинства

методов. По индексу производительности TU полученный метод превосходит большинство методов.

При расчете векторного показателя для каждой категории ВР всех методов полученный метод показывает средний результат: лучший результат получен при расчете показателя для случайных ВР (2-ое место), худший результат получен при расчете показателя для хаотичных ВР (10-ое место).

Анализ полученных результатов прогнозирования показал, что средняя абсолютная ошибка прогноза (MAPE) находится в диапазоне от 0,5% до 4%. Данный показатель является удовлетворительным, т.к. является средним показателем подобной оценки прогноза для известных методов. При этом некоторые методы, показывающие лучшие результаты, не приспособлены для работы с данными в режиме реального времени. Так, например, алгоритм ARIMA (SARIMA) нуждается в предварительной обработке данных для получения высоких результатов прогноза, что неприемлемо для работы в режиме РРВ. При сравнении полученного метода с другими методами прогнозирования по другим критериям (производительность, ошибка) был получен вывод о пригодности алгоритма: алгоритм находится в средней четверке среди двенадцати исследуемых алгоритмов, показывает наилучшие значения индекса производительности TU. Можно сделать вывод, что алгоритм, основанный на использовании сплайна с предварительной настройкой оптимальных параметров, является компромиссом между скоростью и качеством прогноза. Данный подход удовлетворяет принципам работы с данными в режиме реального времени.

## **5. ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ**

Результаты научной деятельности, прежде всего, оцениваются с точки зрения коммерческой ценности проведенной разработки. Данный показатель является одним из основных при поиске источников финансирования научной работы, а также демонстрирует уровень коммерциализации ее результатов. При определении коммерческой ценности разработки учитывают востребованность конечного продукта на рынке, стоимость научной работы, сроки реализации продукта.

Исходя из вышесказанного, целью раздела «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение» является рассмотрение вопросов создания и проектирования такой конкурентноспособной разработки (технологии), способной отвечать современным требованиям в области ресурсоэффективности и ресурсосбережения.

Для достижения поставленной цели в данной работе решаются следующие задачи:

- Оценка коммерческого потенциала, перспективности научного исследования и возможных альтернатив с точки зрения ресурсоэффективности и ресурсосбережения;
- Планирование научного исследования и формирование бюджета работ;
- Определение ресурсной, социальной, финансовой и экономической моделей исследования.

## **5.1. Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения**

### **5.1.1. Потенциальные потребители результатов исследования**

Целевым рынком данной разработки является рынок систем, методов и моделей прогнозирования, работающих с временными рядами. Для подробного анализа потребителей необходимо сегментировать выявленный рынок. Для этого были выделены следующие критерии: уровень подготовки конечного пользователя, вид ресурса для прогнозирования.

Наиболее популярными решениями на данный момент являются библиотеки кода, представляющие собой сборники подпрограмм и объектов для их дальнейшего использования в разработке и исследованиях, и модули, предназначенные для запуска и дальнейшего использования внутри математических пакетов (MatLab, MathCad и т.д.). Помимо этого, в настоящее время широкое распространение получили такие ресурсы как веб-приложения и специализированные десктопные приложения.

Еще один критерий - профессиональные навыки конечного пользователя. Для некоторых современных решений требуется наличие навыка программирования на специализированных языках или навыки работы в распространенных математических решениях. Для других решений – пользователю не обязательно иметь специальных навыков (достаточным будет уверенное владение ПК).

На основании выявленных критериев построена карта сегментирования рынка. Карта представлена в таблице 12. Темным цветом выделены наиболее популярные на данный момент решения. Светлым – существующие, но менее распространенные решения. Решения, выделенные белым – не используются или их использование не имеет смысла.

Таблица 12 – Карта сегментирования рынка

моделей прогнозирования

	Тип предоставляемого ресурса				
		Десктоп-приложение	Библиотека кода	Модуль математического Пакета	Веб-приложение
Профессиональные навыки конечного пользователя	Навыки программирования				
	Навыки работы с математическими Пакетами				
	Отсутствие особых навыков				

Анализ полученной карты сегментирования показал, что наиболее перспективным направлением является разработка десктопных приложений и веб-приложений, так как данный сегмент рынка имеет сравнительно невысокий уровень конкуренции. Подобные решения также востребованы из-за отсутствия специальных требований к навыкам конечных пользователей. Следовательно, данные направления разработки являются привлекательными и перспективными.

### 5.1.2. Анализ конкурентных технических решений

Для анализа были выделены основные существующие аналоги разрабатываемого приложения:

- Библиотека кода Forecast для языка программирования R;
- Программный продукт Caterpillar (веб);
- Математический пакет Neural Networks Toolbox для работы с нейронными сетями в среде MatLab.

Для оценки конкурентоспособности разрабатываемого решения были изучены выделенные решения конкурентов. Результаты анализа представлены в таблице 13.

Таблица 13 – Оценочная карта для сравнения конкурентных технических решений

Критерии оценки	Вес критерия	Баллы				Конкурентоспособность			
		Бф	Бк1	Бк2	Бк3	Кф	Кк1	Кк2	Кк3
<b>Технические критерии оценки ресурсоэффективности</b>									
1. Простота Эксплуатации	0,2	5	2	1	5	0,2	0,08	0,04	0,2
2. Понятность Интерфейса	0,15	5	2	2	5	0,15	0,03	0,03	0,15
3. Функциональные Возможности	0,2	3	5	3	5	0,12	0,2	0,12	0,2
4. Скорость Внедрения	0,05	5	5	4	3	0,05	0,05	0,04	0,03
5. Потребность в ресурсах памяти	0,025	4	4	4	2	0,02	0,02	0,02	0,01
<b>Экономические критерии оценки ресурсоэффективности</b>									
1. Цена	0,15	4	5	4	2	0,12	0,15	0,12	0,06
2. Область Применения	0,15	4	5	3	5	0,12	0,15	0,09	0,15
3. Наличие Обновлений	0,025	5	5	5	5	0,025	0,025	0,025	0,025
4. Уровень проникновения на Рынок	0,05	0	5	4	4	0	0,05	0,04	0,04
Итого:	1					0,805	0,675	0,525	0,865

Результат анализа решений конкурентов показал, что разрабатываемое программное приложение уступает конкурентам в широте функционала. Но широкая линейка функций увеличивает сложность работы с приложением для конечного пользователя (при отсутствии дополнительных навыков). Из этого вытекает преимущество разрабатываемого продукта - простота эксплуатации и понятность интерфейса, что очень высоко оценивается из-за увеличения числа потенциальных пользователей системы. При дальнейшей модернизации приложения функционал может быть расширен.

### 5.1.3. Технология QuaD

Технология QuaD представляет собой инструмент измерения характеристик, которые описывают качество научного исследования и разработки, а также перспективность разработки на рынке, и предназначены для принятия решения целесообразности вложения денежных средств в научно-исследовательский проект.

В соответствии с данной технологией оценке подлежит каждый показатель из группы показателей оценки коммерческого потенциала разработки и группы показателей качества разработки (оценивание по 100-бальной шкале). Веса показателей, определяемые экспертным путем, в сумме должны составлять 1.

Оценка качества и перспективности разрабатываемого решения по технологии QuaD представлена в таблице 14.

Таблица 14 – Оценка качества и перспективности разработки по технологии QuaD

Критерии оценки	Вес критерия	Баллы	Максимальный балл	Относительное значение	Средневзвешенное значение
<b>Показатели оценки качества разработки</b>					
1. Унифицированность	0,075	85	100	0,85	6,375
2. Уровень материалоемкости Разработки	0,075	90	100	0,9	6,75
3. Функциональные Возможности	0,2	75	100	0,75	15
4. Простота Эксплуатации	0,2	100	100	1	20
5. Устойчивость	0,05	95	100	0,95	4,75
<b>Показатели коммерческого потенциала разработки</b>					
1. Конкурентоспособность разработки	0,15	75	100	0,75	11,25
2. Перспективность Рынка	0,15	90	100	0,9	13,5
3. Финансовая эффективность научной разработки	0,075	90	100	0,9	6,75

4. Поддержка продукта	0,025	40	100	0,4	1
Итого	1				85,375

В результате было получено значение, находящееся в диапазоне 80 – 100. Данный факт указывает на допустимую перспективность разработки и качестве проведенного исследования. По результатам оценки качества и перспективности делается вывод об объемах инвестирования в текущую разработки и направлениях ее дальнейшей модернизации.

#### **5.1.4. SWOT-анализ**

SWOT-анализ представляет собой комплексный анализ научного исследования. На первом этапе SWOT-анализа необходимо определить сильные и слабые стороны научного проекта, выявить угрозы и возможности, которое могут проявиться или уже проявились во внешней среде.

К сильным сторонам научной работы относятся особенные, с точки зрения конкуренции, преимущества проекта: ресурсы и возможности, которые могут быть полезны для достижения поставленных целей данной разработки. К слабым сторонам, напротив, относятся особенности проекта, препятствующие достижению поставленной цели.

Возможности представляют собой любую предположительную ситуацию, которая может возникнуть в настоящем или будущем, в условиях внешней среды проекта, которая приведет к улучшению спроса на разрабатываемое решение или к улучшению конкурентоспособности разработки на рынке. К угрозам также относятся возможные в настоящем или будущем ситуации во внешней среде, но влияющие на разработку со знаком минус (спад спроса, угроза конкурентоспособности).

Результаты первого этапа SWOT-анализа представлены в таблице 15.

Таблица 15 – Промежуточная матрица SWOT

	<b>Положительное влияние</b>	<b>Отрицательное влияние</b>
--	------------------------------	------------------------------

<b>Внутренняя среда предприятия</b>	<i>Сильные стороны научного исследования:</i> С1. Понятный и «дружественный» пользовательский интерфейс С2. Простота эксплуатации С3. Небольшие затраты ресурсов ОС	<i>Слабые стороны научного исследования:</i> Сл1. Ограниченность функциональных возможностей системы Сл2. Невозможность пользовательской настройки интерфейса Сл3. Незнание используемой модели прогнозирования на целевом рынке
<b>Внешняя среда предприятия</b>	<i>Возможности научного исследования:</i> В1. Заинтересованность в разработке смежных сегментов рынка В2. Использование разработки в образовательных целях В3. Увеличение популярности прогнозирования в различных сферах производства	<i>Угрозы научного исследования:</i> У1. Появление систем-конкурентов У2. Появление новых моделей прогнозирования с более высоким уровнем точности У3. Потеря интереса клиентов к десктоп-приложениям

На следующем этапе анализа, после того, как для всех областей SWOT-анализа сформулированы пункты, выявляются соответствия сильных и слабых сторон разрабатываемого решения внешним условиям окружающей среды. На данном этапе определяется степень необходимости проведения стратегических изменений для стабилизации влияния внешней среды.

Результатами второго этапа являются интерактивные матрицы проекта, представленные в таблицах 16-19.

Таблица 16 – Интерактивная матрица проекта

<b>Сильные стороны научного исследования</b>				
<b>Возможности научного исследования</b>		C1	C2	C3
	V1	+	+	0
	V2	+	+	+
	V3	+	+	0

Таблица 17 – Интерактивная матрица проекта

<b>Сильные стороны научного исследования</b>				
		C1	C2	C3

научного исследования	У1	+	+	0
	У2	-	-	-
	У3	+	0	-

Таблица 18 – Интерактивная матрица проекта

Слабые стороны научного исследования				
Возможности научного исследования		Сл1	Сл2	Сл3
	В1	+	+	0
	В2	0	0	+
	В3	-	+	+

Таблица 19 – Интерактивная матрица проекта

Слабые стороны научного исследования				
Угрозы научного исследования		Сл1	Сл2	Сл3
	У1	+	+	+
	У2	-	-	0
	У3	-	-	-

На заключительном этапе анализа должна составляется итоговая матрица SWOT-анализа. Конечные результаты учитываются при разработке структуры работ, выполняемых в рамках научного исследования.

Итоговая матрица представлена в приложении А.

### 5.1.5. Определение возможных альтернатив проведения научных исследований

Для определения возможных альтернатив проведения научных исследований рекомендуется использовать морфологический подход, который основан на систематическом исследовании всех возможных в теории вариантов, возникающих из морфологии объекта исследования.

При анализе важных морфологических характеристик объекта исследования, были выделены следующие:

- язык программирования;
- графический интерфейс пользователя;
- вид визуализации;
- формат сохранения данных для анализа;
- способ оптимизации основных параметров модели

прогнозирования.

Для раскрытия всех вариантов для вышеперечисленных характеристик была составлена морфологическая матрица (таблица 20).

Таблица 20 – Морфологическая матрица проекта

	1	2	3
А. Язык программирования	C#	Python	Java
Б. Графический интерфейс Пользователя	Десктоп-приложение	Веб-приложение	
В. Вид визуализации	График «Ящик с усами»	График «Слайн» по временному ряду	Табличный вид
Г. Формат сохранения данных для анализа	Формат .xlsx	Формат .txt	Формат .csv
Д. Способ оптимизации параметров модели Прогнозирования	Нейронные сети	Метаэвристические алгоритмы	Кластеризация

Из данной морфологической матрицы можно выделить три варианта реализации проекта:

Исполнение 1: А1Б1В2Г2Д3

Исполнение 2: А2Б2В2С2Д1

Исполнение 3: А3Б1В1Г1Д2

## 5.2. Планирование и формирование бюджета научных исследований

### 5.2.1. Структура работ в рамках научного исследования

При планировании комплекса предполагаемых работ был разработан перечень этапов и работ в рамках проведения научно-исследовательского проекта. В данном проекте исполнителями являются студент и научный руководитель. Перечень этапов, работ и распределение исполнителей по данным видам работ в рамках проводимого научно-исследовательского проекта представлен в приложении Б.

### 5.2.2. Определение трудоемкости выполнения работ

Для определения стоимости разработки прежде всего необходимо определить трудовые затраты участников научного проекта. Для этого необходимо определить трудоемкость работ каждого из участников проекта.

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

$$t_{ожi} = \frac{3t_{\min i} + 2t_{\max i}}{5}$$

где  $t_{ожi}$  – ожидаемая трудоемкость выполнения  $i$ -ой работы в человеко-днях,  $t_{\min i}$  – минимально возможная трудоемкость выполнения заданной  $i$ -ой работы в человеко-днях, а  $t_{\max i}$  – максимально возможная трудоемкость выполнения заданной  $i$ -ой работы в человеко-днях.

На основе ожидаемой трудоемкости работ, можно определить продолжительность каждой работы в рабочих днях, учитывая, при этом, параллельность выполнения работ несколькими исполнителями. Для этого используется следующая формула:

$$T_{pi} = \frac{t_{ожi}}{Ч_i}$$

где  $T_{pi}$  – продолжительность одной работы в рабочих днях,  $Ч_i$  – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе.

Для удобства построения графика, длительность каждого из этапов работ из рабочих дней следует перевести в календарные дни. Для этого необходимо воспользоваться следующей формулой:

$$T_{ki} = T_{pi} \cdot k_{кал}$$

где  $T_{ki}$  – продолжительность выполнения  $i$ -ой работы в календарных днях,  $T_{pi}$  – продолжительность выполнения  $i$ -ой работы в рабочих днях,  $k_{кал}$  – коэффициент календарности.

Коэффициент календарности определяется по следующей формуле:

$$k_{кал} = \frac{T_{кал}}{T_{кал} - T_{вых} - T_{пр}}$$

где  $T_{кал}$  – количество календарных дней в году,  $T_{вых}$  – количество выходных дней в году,  $T_{пр}$  – количество праздничных дней в году.

Коэффициент календарности в 2020 году – 1,4817.

Все расчеты по данным формулам представлены в приложении В.

### 5.2.3. Разработка графика проведения научного исследования

Одним из способов для представления графика проведения научного исследования является горизонтальный ленточный график - диаграмма Ганта. На данной диаграмме все работы проекта представляются протяженными во времени отрезками с датами начала и окончания выполнения данных этапов проекта.

Диаграмма Ганта с разбивкой по месяцам и декадам за период времени работ по научному проекту представлена в таблице 21.

Таблица 21 – Календарный план-график проведения работ

№	Вид работ	Исполнители	$T_{ki}$	Продолжительность выполнения работ														
				январь		февраль			Март			апрель			май			
				2	3	1	2	3	1	2	3	1	2	3	1	2		
1	Разработка технического Задания	Научный руководитель Ль	1															
2	Выбор направления исследований	Научный руководитель, студент	16															
3	Теоретические и экспериментальные исследования	Студент	77															
4	Обобщение и	Научный	7															



$$Z_m = (1 + k_T) \sum_{i=1}^m C_i \cdot N_{расхi}$$

где  $m$  – количество материальных ресурсов  $i$ -го вида, планируемых к использованию при выполнении научного исследования,  $C_i$  – цена приобретения единицы  $i$ -го вида потребляемых ресурсов,  $k_T$  – коэффициент, учитывающий транспортно-заготовительные расходы.

Материальные затраты, необходимые для разработки программного решения, занесены в таблицу 22.

Таблица 22 – Материальные затраты

Наименование	Единица измерения	Количество			Цена за ед., руб.			Затраты на материалы, руб		
		И1	И2	И3	И1	И2	И3	И1	И2	И3
Бумага, формат А4	Лист	150	150	150	0,5	0,5	0,5	75	75	75
Брюшорование	Штука	1	1	1	100	100	100	100	100	100
Итого								175	175	175

#### 5.2.4.2. Основная заработная плата исполнителей темы

В данную статью бюджета включается основная заработная плата научных работников, непосредственно участвующих в выполнении работ по данной теме. Величина расходов по данной статье определяется исходя из трудоемкости выполняемых работ и действующей системы окладов и тарифных ставок.

Статья включает основную заработную плату работников, непосредственно занятых выполнением научно-техническим исследованием (включая премии и доплаты) и дополнительную заработную плату (12-20 % от основной заработной платы):

$$Z_{зп} = Z_{осн} + Z_{доп}$$

Основная заработная плата рассчитывается по следующей формуле:

$$Z_{осн} = Z_{он} \cdot T_p$$

где  $T_p$  – продолжительность работ, выполняемых научно-техническим работником,  $Z_{дн}$  – среднедневная заработная плата работника, которая рассчитывается по формуле:

$$Z_{дн} = \frac{Z_m \cdot M}{F_d}$$

где  $Z_m$  – месячный должностной оклад работника,  $M$  – количество месяцев работы без отпуска в течение года,  $F_d$  – действительный годовой фонд рабочего времени научно-технического персонала. Баланс рабочего времени приведен в таблице 23.

Таблица 23 – Баланс рабочего времени

Показатели рабочего времени	Руководитель	Студент
Календарное число дней	366	366
Количество нерабочих дней - выходные дни - праздничные дни	118	118
Потери рабочего времени - Отпуск - Невыходы по болезни	24	24
Действительный годовой фонд рабочего времени	224	224

Месячный должностной оклад работника:

$$Z_m = Z_{тс} (1 + k_{пр} + k_{д}) k_p$$

где  $Z_{тс}$  – заработная плата по тарифной ставке, руб;  $k_{пр}$  – премиальный коэффициент (равный 30 % от  $Z_{тс}$ ),  $k_{д}$  – коэффициент доплат и надбавок (примерно 0,2-0,5),  $k_p$  – районный коэффициент (для Томска – 1,3).

Таблица 24 – Расчет основной заработной платы

Исполнители	Разряд	Оклад	$k_{пр}$	$k_p$	$Z_m$	$Z_{дн}$	$T_p$	$Z_{осн}$
Руководитель	4-ый	16819,03	0,3	0,5	39356,53	1293,91	4,8	6210,78
Студент	1-ый	9893,56	0,3	0,2	19291,35	634,23	70,6	44777,07

Размер основной заработной платы для других исполнений приведен в таблице 25.

#### 5.2.4.3. Дополнительная заработная плата исполнителей темы

Расчет дополнительной заработной платы ведется по следующей формуле:

$$Z_{доп} = k_{доп} \cdot Z_{осн}$$

где  $k_{доп}$  – коэффициент дополнительной заработной платы (на стадии проектирования принимается равным 0,12 – 0,15).

#### 5.2.4.4. Отчисления во внебюджетные фонды (страховые отчисления)

В данной статье расходов отражаются обязательные отчисления по установленным законодательством Российской Федерации нормам органам государственного социального страхования (ФСС), пенсионного фонда (ПФ) и медицинского страхования (ФФОМС) от затрат на оплату труда работников. Величина отчислений во внебюджетные фонды определяется исходя из следующей формулы:

$$Z_{внеб} = k_{внеб} (Z_{осн} + Z_{доп})$$

где  $k_{внеб}$  – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.). Для образовательных учреждений это значение равно 27,1 %.

Отчисления во внебюджетные фонды представлены в таблице 25.

Таблица 25 – Отчисления во внебюджетные фонды

Исполнитель	Основная заработная плата			Дополнительная заработная плата		
	И1	И2	И3	И1	И2	И3
Руководитель Проекта	6210,78	6210,78	6210,78	745,29	745,29	745,29
Студент-Дипломник	44777,07	50738,4	50738,4	5373,25	6088,61	6088,61

Коэффициент отчислений во внебюджетные фонды	27,1 %
<b>Итого</b>	
Исполнение 1	72582,05
Исполнение 2	81068,2
Исполнение 3	81068,2

#### 5.2.4.5. Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: оплата услуг связи и интернета, печать и ксерокопирование материалов исследования, оплата электроэнергии, почтовые расходы, размножение материалов и т.д. Их величина определяется по следующей формуле:

$$Z_{\text{накл}} = \left( \frac{\text{сумма\_статей}}{\text{кол-во\_статей}} \right) \cdot k_{\text{нр}}$$

где  $k_{\text{нр}}$  – коэффициент, учитывающий накладные расходы (величина коэффициента накладных расходов – 16%).

#### 5.2.4.6. Формирование бюджета затрат научно-исследовательского проекта

Определение бюджета затрат на научно-исследовательский проект по каждому варианту исполнения приведено в таблице 26.

Таблица 26– Расчет бюджет затрат НТИ

Наименование статьи	Сумма, руб		
	И1	И2	И3
1. Материальные затраты НТИ	175	175	175
2. Затраты на специальное оборудование для научных работ	0	0	0
3. Затраты по основной заработной плате исполнителей темы	50987,85	56949,18	56949,18
4. Затраты по дополнительной заработной плате исполнителей темы	6118,54	6833,9	6833,9
5. Отчисления во внебюджетные фонды	15475,74	17285,12	17285,12
6. Затраты на научные и производственные Командировки	0	0	0
7. Контрагентские расходы	0	0	0

8. Накладные расходы	2690,01	3249,728	3249,728
9. Бюджет затрат НИИ	69940,14	84492,7	84492,7

### 5.3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

Для определения эффективности используется расчет интегральных показателей эффективности научно-исследовательского проекта.

Интегральный показатель финансовой эффективности получают при оценке бюджета затрат нескольких вариантов использования. Для его расчета наибольший интегральный показатель реализации технической задачи принимается за базу расчета, с ним соотносятся финансовые значения по остальным вариантам исполнения.

Интегральный финансовый показатель разработки можно определить по формуле:

$$I_{финр}^{испi} = \frac{\Phi_{pi}}{\Phi_{max}}$$

где  $I_{финр}^{испi}$  – интегральный финансовый показатель разработки,  $\Phi_{pi}$  – стоимость  $i$ -го варианта исполнения,  $\Phi_{max}$  – максимальная стоимость исполнения научно-исследовательского проекта.

Интегральный показатель ресурсоэффективности вариантов исполнения объекта можно определить по следующей формуле:

$$I_{pi} = \sum a_i b_i$$

Где  $I_{pi}$  – интегральный показатель ресурсоэффективности для  $i$ -го варианта исполнения разработки;  $a_i$  – весовой коэффициент  $i$ -го варианта исполнения разработки;  $b_i$  – балльная оценка  $i$ -го варианта исполнения разработки, устанавливается экспертным путем по выбранной шкале оценивания.

Таблица 27 – Сравнительная оценка характеристик вариантов исполнения проекта

Критерии	Весовой коэффициент параметра	И1	И2	И3
1. Способствует росту производительности труда пользователя	0,1	5	5	5
2. Удобство в эксплуатации (соответствует требованиям потребителя)	0,15	5	5	3
3. Помехоустойчивость	0,15	4	3	4
4. Энергосбережение	0,20	4	4	4
5. Надежность	0,25	5	4	5
6. Материалоемкость	0,15	4	4	4
Итого:	1	4,5	4,1	4,2

Интегральный показатель эффективности вариантов использования разработки определяется на основании интегрального показателя по формуле:

$$I_{\text{исп}i} = \frac{I_{\text{р-исп}i}}{I_{\text{финр}}}$$

Сравнительная эффективность проекта вычисляется как отношение интегральных показателей эффективности вариантов исполнения разработки.

Вычисления, необходимые для определения сравнительной эффективности разработки для поиска наиболее эффективного решения задачи, представленной в бакалаврской работе, приведены в таблице 28.

Таблица 28 – Сравнительная эффективность разработки

№ п/п	Показатели	Исполнение 1	Исполнение 2	Исполнение 3
1	Интегральный финансовый показатель разработки	0,828	1	1
2	Интегральный показатель ресурсоэффективности Разработки	4,5	4,1	4,1
3	Интегральный показатель Эффективности	5,435	4,1	4,1
4	Сравнительная эффективность вариантов Исполнения	1	0,754	0,754

Исходя из результатов сравнительной эффективности разработки наиболее эффективным решением с позиции финансовой и ресурсной эффективности является первый вариант исполнения.

## 6. СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ

Целью данной выпускной квалификационной работы является разработка десктопного приложения, предназначенного для краткосрочного прогнозирования на основе штрафных р-сплайнов. Настройка одного из основных параметров, используемого математического аппарата происходит при помощи алгоритма кластерного анализа DBSCAN. Разработанное приложение позволит пользователю загрузить необходимые данные, выставить самостоятельно требуемые для работы приложения параметры и получить краткосрочный прогноз по этим данным. Результат работы приложения будет представлен пользователю графически: исходные данные с разделением на кластеры, прогнозируемые значения. Также для пользователя доступна функция экспорта табличного представления результатов в некоторые форматы.

Основными потенциальными областями применения для разработанного приложения являются области исследовательской и образовательной деятельности, а также финансовые, промышленные, энергетические и другие предприятия, нуждающиеся в возможности краткосрочного прогнозирования в масштабе реального времени.

Десктопное приложение будет использоваться на рабочем месте оператора ПК, следовательно, в данном разделе работы будут рассмотрены вредные и опасные факторы, которые могут возникнуть при использовании средств вычислительной техники. Также данный раздел содержит выявление возможных вредных воздействий на окружающую среду, программ по их снижению и экономии невозполнимых ресурсов и способах защиты в ЧС, которые могут возникнуть на рабочем месте оператора.

## **6.1. Правовые и организационные вопросы обеспечения безопасности**

### **6.1.1. Специальные правовые нормы трудового законодательства**

Регулирование отношений между работником и работодателем, касающихся оплаты труда, трудового распорядка, особенности регулирования труда женщин, детей, людей с ограниченными способностями и проч., осуществляется законодательством РФ, а именно трудовым кодексом РФ.

Режим труда и отдыха предусматривает соблюдение определенной длительности непрерывной работы на персональном компьютере (ПК) и перерывов, регламентированных с учетом продолжительности рабочей смены, видов и категории трудовой деятельности. Вид трудовой деятельности на персональном компьютере при работе с разработанным приложением соответствует группе «В» (творческая работа в режиме диалога с ПК), категория трудовой деятельности – III (до 6 часов непосредственной работы на ПК). При 8-часовой рабочей смене и работе на ПК, соответствующей описанным выше критериям, необходимо выполнять один из видов дополнительных перерывов: через 1,5- 2,0 часа от начала рабочей смены и конца обеденного перерыва - регламентированные перерывы в 20 минут или перерыв в 15 минут через каждый час работы.

Продолжительность непрерывной работы на ПК без регламентированного перерыва не должна превышать 2 часа. Эффективными являются нерегламентированные перерывы (микропаузы) длительностью 1-3 минуты. Во время перерывов (любого вида) целесообразно выполнять какой-либо комплекс упражнений, направленный на поддержание здорового функционирования организма: гимнастика для глаз, пальцев рук, а также массаж. Данные комплексы упражнений необходимо обновлять раз в 2-4 недели. Продолжительность рабочего дня не должна быть меньше указанного времени в договоре, но не больше 40 часов в неделю. Для работников до 16 лет

– не более 24 часов в неделю, от 16 до 18 лет и инвалидов I и II группы – не более 35 часов. При работе в ночное время продолжительность рабочей смены сокращается на один час. К работе в ночную смену не допускаются беременные женщины, работники, не достигшие возраста 18 лет, женщины, имеющие детей в возрасте до трех лет, инвалиды, работники, имеющие детей и инвалидов, а также работники, осуществляющие уход за больными членами их семей в соответствии с медицинским заключением, матери и отцы-одиночки детей до пяти лет.

Организация обязана предоставлять ежегодный отпуск продолжительностью 28 календарных дней. Дополнительные отпуска предоставляются работникам, занятым на работах с вредными или опасными условиями труда, работникам имеющими особый характер работы, работникам с ненормированным рабочим днем и работающим в условиях Крайнего Севера и приравненных к нему местностях.

### **6.1.2. Организационные мероприятия при компоновке рабочей зоны**

Рабочее место должно быть организовано в соответствии с требованиями стандартов, технических условий и (или) методических указаний по безопасности труда. Оно должно удовлетворять следующим требованиям:

- обеспечивать возможность удобного выполнения работ;
- учитывать физическую тяжесть работ;
- учитывать размеры рабочей зоны и необходимость передвижения в ней работающего;
- учитывать технологические особенности процесса выполнения работ.

Невыполнение требований к расположению и компоновке рабочего места может привести к получению работником производственной травмы или развития у него профессионального заболевания. Рабочее место инженера-программиста должно соответствовать требованиям СанПин 2.2.2/2.4.1340-03.

Конструкция оборудования и рабочего места при выполнении работ в положении сидя должна обеспечивать оптимальное положение работающего, которое достигается регулированием высоты рабочей поверхности, высоты сидения, оборудованием пространства для размещения ног и высотой подставки для ног. Схемы размещения рабочих мест с персональными компьютерами должны учитывать множество факторов. Расстояние между боковыми поверхностями мониторов - не менее 1,2 м, расстояние между экраном монитора и тыльной частью другого монитора - не менее 2,0 м. Рекомендуется изолировать рабочие места друг от друга перегородками высотой 1,5 - 2,0 м. Клавиатура должна располагаться на поверхности стола на расстоянии 100-300 мм от края, обращенного к пользователю.

## **6.2. Производственная безопасность**

Чтобы обеспечить производственную безопасность человека на рабочем месте, необходимо проанализировать возможное воздействие вредных и опасных факторов, которые могут возникнуть при разработке и использовании научно-исследовательского проекта. Производственные факторы считаются вредными, если их воздействие на работающего в определенных условиях приводит к заболеванию или снижению трудоспособности. Заболевания, возникающие под действием вредных производственных факторов, называются профессиональными. Производственные факторы относятся к опасным, если их воздействия на работающего в определенных условиях приводит к травме или к другому внезапному резкому ухудшению здоровья. Воздействие вредных факторов приводит к снижению работоспособности, появлению и развитию утомления, связанного с изменениями, возникающими в работе центральной нервной системы. При этом вредные факторы проявляются при условии интенсивности и длительности воздействия. В отличие от них, опасные факторы оказывают моментальное влияние на здоровье работника, они могут привести к травмам (ожогам, облучению, отравлению и т.д.). В таблицах 29-30

предоставлены возможные вредные и опасные факторы, которые могут возникнуть при работе инженера-программиста с разработанным приложением.

Таблица 29 – Вредные факторы, возникающие при работе с разработанным приложением для краткосрочного прогнозирования

Наименование видов работ	Факторы по ГОСТ 12.0.003-2015	Нормативные документы
Работа за персональным компьютером (ПК)	Отклонение показателей микроклимата (температуры и влажности воздуха)	СанПиН 2.2.2/2.4.1340-03 [2] СанПиН 2.2.4.548-96 [3]
	Недостаточная освещенность рабочей зоны	СанПиН 2.2.2/2.4.1340-03

Таблица 30 – Опасные факторы, возникающие при работе с разработанным приложением для краткосрочного прогнозирования

Наименование видов работ	Факторы по ГОСТ 12.0.003-2015	Нормативные документы
Работа за персональным компьютером (ПК)	Опасность поражения электрическим током	ГОСТ 12.1.038–82 [4]

## **6.2.1. Анализ вредных и опасных факторов**

### **6.2.1.1. Отклонение показателей микроклимата в помещении**

Микроклимат в помещении является одним из самых необходимых для обеспечения благоприятных условий труда для работников производственным фактором, так как он оказывает большое влияние на тепловое самочувствие человека. Микроклимат в производственном помещении, прежде всего, зависит от внешних условий таких, как категория работ, период года, условия вентиляции, а также от особенностей самого технологического процесса. Можно выделить следующие параметры, характеризующие микроклимат в производственных помещениях: температура воздуха ( $t$ , °C), температура поверхностей ( $t$ , °C), относительная влажность воздуха ( $\phi$ , %), скорость движения воздуха ( $v$ , м/с), интенсивность теплового облучения ( $I$ , Вт/м<sup>2</sup>).

В производственных помещениях при работе с персональными вычислительными машинами происходит постоянное выделение тепла вычислительной техникой, средствами освещения и другими

вспомогательными приборами. И так как оператор ПК расположен в непосредственной близости с источниками выделения тепла, то высокая температура воздуха может способствовать быстрому перегреву программиста и быстрой утомляемости [5]. Помимо температуры воздуха, большое влияние на терморегуляцию организма оказывает относительная влажность воздуха. Как отмечено в [6] высокие показатели влажности (значения больше 85 %) затрудняют терморегуляцию тела и снижают возможность испарения пота, низкие показатели влажности (значения менее 20 %) вызывают пересыхание слизистых оболочек. Работа инженера-программиста относится к категории Ia (работа производится сидя и сопровождается незначительным физическим напряжением, интенсивность энергозатрат до 139 Вт). Оптимальные значения показателей микроклимата на рабочих местах производственных помещений согласно СанПиН 2.2.4.548-96 для категории работ Ia представлены в таблице 31. [3]

Таблица 31 – Оптимальные величины показателей микроклимата на рабочих местах производственных помещений

Период года	Категория работ	Температура воздуха, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	Ia	22 – 24	40 – 60	0,1
Теплый	Ia	23 – 25	40 – 60	0,1

Допустимые микроклиматические условия не вызывают повреждений или нарушений состояния здоровья, но могут приводить к возникновению общих и локальных ощущений теплового дискомфорта, напряжению механизмов терморегуляции, ухудшению самочувствия и понижению работоспособности. Допустимые величины показателей микроклимата на рабочих местах согласно СанПиН 2.2.4.548-96 приведены в таблице 32 (для соответствующей категории). [3]

Таблица 32 – Допустимые величины показателей микроклимата на рабочих местах производственных помещений

Период года	Категория работ	Температура воздуха, °С		Относительная влажность воздуха, %	Скорость движения воздуха, м/с	
		Диапазон ниже оптимальных величин	Диапазон ниже оптимальных величин		Для диапазона температур воздуха ниже оптимальных величин, не более	Для диапазона температур воздуха ниже оптимальных величин, не более
Холодный	Ia	20,0 – 21,9	24,1 – 25,0	15 – 75	0,1	0,1
Теплый	Ia	21,0 – 22,9	25,1 – 28,0	15 – 75	0,1	0,2

Согласно требованиям СанПиН 2.2.2/2.4.1340–03, в кабинете поддерживается температура равная 19–20 С°, при относительной влажности в 55–58%. [2] Для поддержания данных условий микроклимата необходимо проводить в помещении ежедневную влажную уборку и систематически проветривать помещение.

### 6.2.1.2. Недостаточная освещенность рабочей зоны

Недостаточная освещенность рабочей зоны – вредный производственный фактор, который регламентируется СП 52.13330.2011. Недостаточный уровень освещенности в помещении приводит к снижению остроты зрения, головным болям, снижению концентрации внимания и к ухудшению производительности труда в целом. Причины недостаточной освещенности: недостаточность естественного освещения, недостаточность искусственного освещения, пониженная контрастность.

В рабочем помещении должны присутствовать естественное и искусственное освещение. Коэффициент естественного освещения должен быть не менее 1,2%. Согласно СанПиН 2.2.2/2.4.1340-03 освещенность на поверхности рабочего 55 стола в зоне размещения документа должна быть 300 – 500 лк. Данное требование может быть достигнуто путем установки местного

небликующего освещения. Освещенность поверхности экрана не должна превышать 300 лк. Яркость светящихся поверхностей (окон, светильников), находящихся в поле зрения должна быть не более 200 кд/м<sup>2</sup>. Для источников искусственного освещения следует применять люминесцентные лампы типа ЛБ и компактные люминесцентные лампы (КЛЛ). Коэффициент пульсации при работе с компьютером не должен превышать 5%. Помимо этого необходимо ограничить отраженную блескость на рабочих поверхностях (экран, рабочий стол и т.д.) за счет правильного выбора и расположения светильников. Яркость бликов на экране ПК не должна превышать 40 кд/м<sup>2</sup>. Светильники местного освещения должны иметь непросвечивающий отражатель.

### **6.2.1.3. Опасность поражения электрическим током**

Большая часть работы оператора ПК проходит за электрооборудованием. Именно поэтому необходимо уделять большое внимание электробезопасности.

Опасность поражения человека электрическим током оценивается величиной тока  $I$  (А), проходящего через его тело, или напряжением прикосновения  $U$  (В). Степень опасного воздействия на человека электрического тока зависит от рода и величины напряжения тока, частоты электрического тока, пути тока через тело человека, продолжительности его воздействия на организм человека, а также условий внешней среды. Электрический ток, протекая через тело человека, производит термическое, механическое и световое воздействие – электролитическое разложение жидкости (в том числе и крови), судорожное сокращение мышц, разрыв тканей и поражение глаз.

Поражение электрическим током при работе с компьютерами возможно при наличии оголенных участков на кабеле, нарушении изоляции распределительных устройств и от токоведущих частей компьютера в случае их пробоя и нарушении изоляции, при контакте электронно-вычислительной машины с влажными поверхностями (руки, одежда). Помещение, где

располагается рабочее место оператора ПК, можно отнести к помещениям без повышенной опасности, так как в нем отсутствуют следующие факторы: сырость, токопроводящая пыль, токопроводящие полы, высокая температура, возможность одновременного прикосновения человека к имеющим соединение с землей металлоконструкциям зданий, технологическим аппаратам, механизмам и металлическим корпусам электрооборудования. Для предотвращения возможности поражения электрическим током можно проводится целый ряд мероприятий (использование только исправных инструментов, аттестованных службой КИПиА, заземление приборов, поддержание исправности электропроводки и т.д.).

[2] Согласно ГОСТ 12.1.038-82 на рабочем месте инженера-программиста допускаются уровни напряжений прикосновения и токов, представленные в таблице 33. [4]

Таблица 33 – Предельно допустимые напряжения прикосновения и токи

Род тока	Напряжение прикосновения, В	Ток, мА
Переменный, 50 Гц	Не более 2,0	Не более 0,3
Постоянный	Не более 8,0 Не	более 1,0

### **6.3. Экологическая безопасность**

#### **6.3.1. Анализ воздействия продукта на окружающую среду**

Постоянное развитие научно-технического прогресса приводит к негативному воздействию на окружающую среду. Увеличение количества компьютерных систем, внедряемых в производственную сферу, приводит к увеличению объема потребляемой электроэнергии, что влечет за собой увеличение мощностей электростанций и их количества. И то, и другое содействует нарушению экологической обстановки. При этом выбросы со станций оказывают негативное влияние на атмосферу.

Основным фактором, оказывающим негативное влияние на гидросферу и литосферу при работе с реализованным приложением, является образование

отходов. В помещении образуются следующие виды отходов: макулатура, отходы от продуктов питания и личной гигиены (упаковка, органические отходы), отходы от канцелярских принадлежностей, отходы от офисной техники (использованные картриджи, упаковка, неисправные компоненты), лампы.

### **6.3.2. Решения по обеспечению экологической безопасности**

Наиболее активной формой защиты окружающей среды от вредного воздействия выбросов промышленных предприятий является полный переход к безотходным и малоотходным технологиям и производствам. Это потребует решения целого комплекса сложных технологических, конструкторских и организационных задач, основанных на использовании новейших научно-технических разработках. Следует использовать современные ЭВМ с режимом пониженного потребления электроэнергии при длительном простое для снижения энергопотребления. Для каждого вида отходов должны применяться свои методы переработки и утилизации. Все отходы следует собирать, сортировать и направлять на переработку в соответствующие организации.

## **6.4. Безопасность в чрезвычайных ситуациях**

### **6.4.1. Перечень возможных ЧС при разработке и эксплуатации**

Помещение для работы операторов ПЭВМ по системе классификации категорий помещений по взрывопожарной и пожарной опасности относится к категории Д (из категорий А, Б, В1-В4, Г, Д), т.к. относится к помещениям с негорючими веществами и материалами в холодном состоянии [8].

Чрезвычайные ситуации, которые могут возникнуть при разработке и эксплуатации реализованного программного продукта:

- техногенные (взрывы, пожары, обрушение помещений, аварии на системах жизнеобеспечения);
- природные (наводнения, ураганы, бури, природные пожары);
- биологические (эпидемии, пандемии);

- социальные (террор, нарушение границ и политики).

Наиболее характерной для объекта, где размещаются рабочие помещения, оборудованные электронно-вычислительными машинами, чрезвычайной ситуацией является пожар.

Причинами возникновения данного вида ЧС могут являться:

- возникновением короткого замыкания в электропроводке;
- возгоранием устройств вычислительной техники из-за неисправности аппаратуры;
- возгоранием устройств искусственного освещения;
- возгоранием мебели по причине нарушения правил пожарной безопасности, а также неправильного использования дополнительных бытовых электроприборов и электроустановок.

#### **6.4.2. Разработка действий в результате возникшей ЧС и меры по ликвидации ее последствий**

Основной задачей современной пожарной безопасности является пожарная профилактика: устранение непосредственных или возможных причин пожаров, создание безопасных условий для вынужденной эвакуации людей из здания, обеспечение условий для тушения возможного огня и т.д.

Обеспечение пожарной безопасности — приоритетная задача для любого предприятия. Создание системы защиты регламентировано законом и нормативными документами различных ведомств. Каждый сотрудник организации должен быть ознакомлен с инструкцией по пожарной безопасности, пройти инструктаж по технике безопасности и строго соблюдать его. Запрещается использовать электроприборы в условиях, не соответствующих требованиям инструкций изготовителей, или имеющие неисправности, которые в соответствии с инструкцией по эксплуатации могут привести к пожару, а также эксплуатировать электропровода и кабели с поврежденной или потерявшей защитные свойства изоляцией.

Электроустановки и бытовые электроприборы в помещениях по окончании рабочего времени должны быть обесточены (кроме дежурного освещения и пожарной сигнализации).

Недопустимо хранение легковоспламеняющихся, горючих и взрывчатых веществ, использование открытого огня в помещениях офиса. С периодичностью не реже одного раза в три года необходимо проводить замеры сопротивления изоляции токоведущих частей силового и осветительного оборудования.

Для тушения пожара необходимо применять ручные углекислотные огнетушители (типа ОУ-2, ОУ-5), находящиеся в помещениях офиса, и пожарный кран внутреннего противопожарного водопровода. Они предназначены для тушения начальных возгораний различных веществ и материалов, за исключением веществ, горение которых происходит без доступа воздуха. Огнетушители должны постоянно содержаться в исправном состоянии и быть готовыми к действию. Категорически запрещается тушить возгорания в помещениях офиса при помощи химических пенных огнетушителей (типа ОХП-10).

#### **6.5. Выводы по разделу**

При написании раздела «Социальная ответственность» для выпускной квалификационной работы бакалавра был проведен анализ взаимного влияния разрабатываемого программного решения и окружающей среды. Выявлены вредные и опасные производственные факторы, разработаны и описаны мероприятия по их предотвращению. Была описана возможность возникновения некоторых чрезвычайных ситуаций, предложены варианты действий по их предотвращению, обеспечению безопасности и ликвидации последствий ЧС. Данный анализ позволит обеспечить наиболее комфортную работу над исследованием и созданием программного решения, а также позволит предотвратить вредное воздействие на окружающую среду.

## ЗАКЛЮЧЕНИЕ

В результате выполнения выпускной квалификационной работы бакалавра был разработан метод прогнозирования временных рядов на основе штрафного р-сплайна и оптимизации параметров сплайна с помощью кластерного анализа. Данный метод был адаптирован для работы с данными в режиме реального времени.

Полученный метод был сравнен с другими известными методами прогнозирования – результаты сравнения оказались удовлетворительными, получено высокое значение индекса производительности TU. Оценка ошибки прогнозирования показала значения в интервале от 0,5% до 4%, что является удовлетворительным показателем для прогноза.

Для работы с полученным методом было разработано десктопное приложение на языке программирования C# с использованием технологии Window Forms.

Разработанное приложение обеспечивает возможность получить прогноз по пользовательским или модельным данным. Пользователю доступно получение оптимальных параметров для построения прогноза и их изменение вручную. Пользователь может получить графическое представление значений прогноза и оценку результатов прогноза в нескольких вариантах ошибок прогнозирования.

## CONCLUSION

As a result of the bachelor's final qualification work, a method for predicting time series based on a penalty p-spline and optimizing spline parameters using cluster analysis was developed. This method has been adapted to work with real-time data.

The obtained method was compared with other known forecasting methods – the comparison results were satisfactory and a high value of the Theil's U coefficient was obtained. The estimation of the forecast error showed values in the range from 0.5% to 4%, which is a satisfactory indicator for the forecast.

To work with the obtained method, a desktop-application was developed in the C# programming language using the Window Forms technology.

The developed application allows you to get a forecast based on user or model data. The user can get the optimal parameters for making a forecast and change them manually. The user can get a graphical representation of forecast values and an estimate of forecast results in several variants of forecast errors.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Охрана труда [Электронный ресурс] / Безопасность жизнедеятельности. URL:<http://www.grandars.ru/shkola/bezopasnostzhiznedeyatelnosti/ohranatruda.html>, свободный (дата обращения: 10.03.2020).
2. СанПиН 2.2.2/2.4.1340-03 Гигиенические требования к персональным электронно-вычислительным машинам и организации работы.
3. СанПиН 2.2.4.548-96 Гигиенические требования к микроклимату производственных помещений.
4. ГОСТ 12.1.038–82 Система стандартов безопасности труда. Электробезопасность. Предельно допустимые значения напряжений прикосновения и токов.
5. Федеральный закон от 22.07.2008 N 123-ФЗ (ред. От 13.07.2015) «Технический регламент о требованиях пожарной безопасности» [Электронный ресурс]/КонсультантПлюс. URL:[http://www.consultant.ru/document/cons\\_doc\\_LAW\\_78699/](http://www.consultant.ru/document/cons_doc_LAW_78699/), свободный (дата обращения: 13.03.2020).
6. Ефремова О. С. Требования охраны труда при работе на персональных электронно-вычислительных машинах. – 2-е изд., перераб. И доп. – М.: Издательство «Альфа-Пресс», 2008.
7. Назаренко О. Б. Безопасность жизнедеятельности: учебное пособие / О. Б. Назаренко, Ю. А. Амелькович; Томский политехнический университет. – 3-е изд., перераб. и доп. – Томск: Изд-во Томского политехнического университета, 2013.
8. НПБ 105-03 Определение категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности [Электронный ресурс] / Электронный фонд правовой и нормативно-технической документации. URL:<http://docs.cntd.ru/document/1200032102>, свободный (дата обращения: 11.03.2020).
9. Е. А. Кочегурова, Е. С. Горохова. Текущее оценивание производной нестационарного процесса на основе рекуррентного сглаживающего сплайна//

Автометрия. – 2016. - Т. 52. № 3. – [С. 79-85].

10. Плотностный алгоритм кластеризации пространственных данных с присутствием шума — DBSCAN. [Электронный ресурс] / URL: <https://habr.com/post/143151/> / (дата обращения 12.03.2020).

11. Carmela Iorio, Gianluca Frasso, Antonio D'Ambrosio Parsimonious time series clustering using P-splines // Expert Systems with Applications №52, 15 June 2016, Pages 26-38.

12. Antonio Rafael Sabino Parmezan, Vinicius M.A. Souza, Gustavo E.A.P.A. Batista. Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model // Information Sciences №484, May 2019, Pages 302-337.

13. Кочегурова Е. А. Использование вариационного сглаживающего сплайна в задаче краткосрочного прогнозирования / Е. А. Кочегурова, Е. В. Шебеко // Известия Томского политехнического университета [Известия ТПУ]. — 2006. — Т. 309, № 7. — [С. 36-39].

14. Завьялов Ю. С., Квасов Б. И., Мирошниченко В. Л. Методы сплайн-функций. М.: Наука, 1980. 382 с.

15. Аверкин А.Н, Ярушев С.А. Прогнозирование временных рядов на основе гибридных нейронных сетей // Наука и образование: научное издание МГТУ им. Н.Э. Баумана – 2016. - №12. - [С. 233-246].

16. Abdoos Adel Hemmati, Mohammad Abdoos, Ali Akbar. Short term load forecasting using a hybrid intelligent method // Knowledge-Based System №76, Mar 2015, p139-147.

17. Чучуева И. А. Модель прогнозирования временных рядов по выборке максимального подобия [Электронный ресурс] : дис. ... канд. тех. наук / Московский государственный технический университет им. Н. Э. Баумана. Москва, 2012.

18. Низамитдинов А.И. Использование штрафного сплайна (p-spline) и кубического сглаживающего сплайна в прогнозировании временных рядов //

Вестник Таджикского государственного университета права, бизнеса и политики. Серия гуманитарных наук - 2012. - № 4. - [С. 124-131].

19. Создание приложения Windows Forms на C# в Visual Studio./ URL: <https://docs.microsoft.com/ru-ru/visualstudio/ide/create-csharp-winform-visual-studio?view=vs-2019/> (дата обращения: 01.05.2020).

20. UML-диаграммы классов./ URL: <https://prog-cpp.ru/uml-classes/> (дата обращения: 12.05.2020).

21. Классификация методов и моделей прогнозирования [Электронный ресурс]: Хабрахабр // <https://habr.ru> URL: <https://habr.com/ru/post/177633/> (дата обращения: 13.04.2020).

22. Saeed Aghabozorgi, Ali Seyed Shirkhorshidi, TehYing Wah. Time-series clustering – A decade review // Information Systems № 53, October–November 2015, Pages 16-38.

23. Alexey Golubev, Anh Tuan Ngyuen, Maxim Shcherbakov, Tran Van Phu. Clustering helps to determine the changes in telecom subscribers' behavior// Advances in Computer Science Research №72, May 2017, Pages 339-343.

24. Лоскутов А.Ю. Анализ временных рядов. Курс лекций. – М.: Физический факультет МГУ (год неизвестен). – 113 с.

25. Maxim Vladimirovich Shcherbakov, Adriaan Brebels, Nataliya Lvovna Shcherbakova. A Survey of Forecast Error Measures// World Applied Sciences Journal №14, Sept 2013, Pages 171-176.

26. Антоненков, Д. В. Краткосрочное прогнозирование электропотребления угольного разреза в условиях оптового рынка электроэнергии : дис. ... канд. техн. наук : 05.09.03 / Антоненков Дмитрий Васильевич. - Красноярск, 2009. - 192 с.

27. Кондратьев М. А. Методы прогнозирования и модели распространения заболеваний// Компьютерные исследования и моделирование - 2013. - Т.5 №5. - [С. 863–882].

28. Ильясов Р.Х. Сплайн-технологии моделирования, анализа и прогнозирования динамики экономических процессов при наличии сезонности//Актуальные вопросы современной науки-2008. - №3. - [С. 380-391].
29. L. Ye, C. Qiuru, X. Naixu, L. Yijun и Z. Guangping. Customer Segmentation for Telecom with the k-means Clustering// Information Technology Journal № 3, 2013, Pages 409–413.
30. Ермолаев, В.А. О методах прогнозирования временных рядов и непрерывных процессов// Радиотехнические и телекоммуникационные системы. - 2016. - №2. - [С. 52-63].
31. Маевский В.В., Харин Ю.С. Робастность регрессионного прогнозирования при наличии функциональных искажений модели // Автоматика и телемеханика. - 2012. - №11. - [С. 118-137].
32. Вьюгин В. В., Трунов В. Г. Адаптивное прогнозирование и его применение для технического анализа финансовых инструментов// Информационные процессы. - 2011. - Т. 11, № 1. – [С. 46–75].

## ПРИЛОЖЕНИЕ А

### Матрица SWOT-анализа

	<p><b>Сильные стороны научного исследования:</b>  С1. Понятный и «дружественный» пользовательский интерфейс  С2. Простота эксплуатации  С3. Небольшие затраты ресурсов ОС</p>	<p><b>Слабые стороны научного исследования:</b>  Сл1. Ограниченность функциональных возможностей системы  Сл2. Невозможность пользовательской настройки интерфейса  Сл3. Неизвестность используемой модели прогнозирования на целевом рынке</p>
<p><b>Возможности научного исследования:</b>  В1. Заинтересованность в разработке смежных сегментов рынка  В2. Использование разработки в образовательных целях  В3. Увеличение популярности прогнозирования в различных сферах производства.</p>	<p><b>Направления развития:</b>  1. В1В2В3С1С2 – модернизация приложения, визуализация результатов, отсутствие «минимального порога» вхождения  2. В2С3 – оптимизация приложения для уменьшения затрат потребляемой памяти и ресурсов операционной системы</p>	<p><b>Сдерживающие факторы:</b>  1. В1Сл1Сл2 – невозможность кастомизации приложения под нужды конкретного пользователя  2. В2Сл3 – отсутствие популярных научных исследований используемой модели прогнозирования  3. В3Сл2Сл3 – отсутствие популярности разработки на рынке</p>
<p><b>Угрозы научного исследования:</b>  У1. Появление систем-конкурентов  У2. Появление новых моделей прогнозирования с более высоким уровнем точности  У3. Потеря интереса клиентов к десктоп-приложениям</p>	<p><b>Угрозы развития:</b>  1. В1Сл1Сл2 – отсутствие интереса представителей смежных областей к данному приложению  2. В3Сл2Сл3 – Возникновение интереса к другим видам прогнозирования и другим типам приложений</p>	<p><b>Уязвимости:</b>  1. У1Сл1Сл2Сл3 – отсутствие перспективных преимуществ перед компаниями-конкурентами</p>

## ПРИЛОЖЕНИЕ Б

### Перечень этапов, работ и распределение исполнителей

Основные этапы	№ раб	Содержание работ	Должность исполнителя
1	2	3	4
Разработка технического задания	1	Составление и утверждение технического задания	Научный руководитель
Выбор направления исследований	2	Принятие решения о разработке, постановка целей и задач, получение исходных данных	Научный руководитель
	3	Подбор и изучение материалов по Тематике	Студент
	4	Календарное планирование работ по теме	Научный руководитель, студент
Теоретические и экспериментальные исследования	5	Изучение существующих технологий краткосрочного прогнозирования	Студент
	6	Изучение используемой в разработке модели прогнозирования	Студент
	7	Определение структуры разрабатываемого ПО, выбор средств разработки	Студент
	8	Работа над отдельными модулями Программы	Студент
	9	Редактирование и отладка Программы	Студент
Обобщение и оценка результатов	10	Анализ результатов работы	Научный руководитель, Студент
	11	Выбор метода оптимизации входных параметров	Научный руководитель, студент
	12	Анализ результатов после внедрения оптимизации	Научный руководитель, Студент
Разработка технической документации и проектирование	13	Выпуск документации по программному приложению	Студент
	14	Оформление графического материала, визуализация данных	Студент
	15	Оценка эффективности и применения разработки	Студент
	16	Разработка социальной ответственности по теме	Студент
Оформление отчета	17	Составление пояснительной записки	Студент

## ПРИЛОЖЕНИЕ В

### Временные показатели научного исследования

Название работы	Исполнители	Трудоемкость работ, человеко-дни									Длительность работ					
		Tmin			Tmax			Тожи			Тр, рабочие дни			Тк, календарные дни		
		И1	И2	И3	И1	И2	И3	И1	И2	И3	И1	И2	И3	И1	И2	И3
Составление и утверждение технического задания	1	1	1	1	2	2	2	1,4	1,4	1,4	1,4	1,4	1,4	2	2	2
Принятие решения о разработке, постановка целей и задач, получение исходных данных	1	1	1	1	3	3	3	1,8	1,8	1,8	1,8	1,8	1,8	3	3	3
Подбор и изучение материалов по тематике	1	5	6	6	10	11	11	7	8	8	7	8	8	10	12	12
Календарное планирование работ по теме	2	1	1	1	2	2	2	1,4	1,4	1,4	0,7	0,7	0,7	1	1	1
Изучение существующих технологий краткосрочного Прогнозирования	1	10	10	10	15	15	15	12	12	12	12	12	12	18	18	18
Изучение используемой в разработке модели Прогнозирования	1	7	7	7	10	10	10	8,4	8,4	8,4	8,4	8,4	8,4	12	12	12
Определение структуры разрабатываемого ПО,	1	2	5	3	3	6	4	2,4	5,4	3,4	2,4	5,4	3,4	4	8	5



Составление пояснительной записки (эксплуатационно- технической документации)	1	3	3	3	5	5	5	3,8	3,8	3,8	3,8	3,8	3,8	6	6	6
<b>Итого:</b>								<b>80,2</b>	<b>89,6</b>	<b>89,6</b>	<b>75,4</b>	<b>84,8</b>	<b>84,8</b>	<b>114</b>	<b>128</b>	<b>128</b>