

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Школа Информационных технологий и робототехники
Направление подготовки 09.04.02 «Информационные системы и технологии»
Отделение школы (НОЦ) Информационных технологий

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Анализ индуктивного метода прогнозирования временных рядов и поиск мета-характеристик для улучшения его репрезентативности

УДК 004.8.519.246.8

Студент

Группа	ФИО	Подпись	Дата
8ИМ6А	Гайнудинов Тимур Сохибназарович		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Буркатовская Ю.Б.	к.ф. – м.н., доцент		

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН	Старикова Е.В.	к.филос.н., доцент		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОКД	Король И.С.	к.х.н., доцент		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ	Марков Н.Г.	д.т.н., профессор		

Томск – 2018 г.

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Код результатов	Результат обучения (выпускник должен быть готов)
<i>Профессиональные и общепрофессиональные компетенции</i>	
P1	Воспринимать и самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте.
P2	Владеть и применять методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе в глобальных компьютерных сетях.
P3	Демонстрировать культуру мышления, способность выстраивать логику рассуждений и высказываний, основанных на интерпретации данных, интегрированных из разных областей науки и техники, выносить суждения на основании неполных данных, анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями.
P4	Анализировать и оценивать уровни своих компетенций в сочетании со способностью и готовностью к саморегулированию дальнейшего образования и профессиональной мобильности. Владеть, по крайней мере, одним из иностранных языков на уровне социального и профессионального общения, применять специальную лексику и профессиональную терминологию языка.
P5	Разрабатывать стратегии и цели проектирования, критерии эффективности и ограничения применимости, новые методы, средства и технологии проектирования геоинформационных систем (ГИС) или промышленного программного обеспечения.
P6	Планировать и проводить теоретические и экспериментальные исследования в области создания интеллектуальных ГИС и ГИС технологии или промышленного программного обеспечения с использованием методов системной инженерии.
P7	Осуществлять авторское сопровождение процессов проектирования, внедрения и сопровождения ГИС и ГИС технологий или промышленного программного обеспечения с использованием методов и средств системной инженерии, осуществлять подготовку и обучение персонала.

P8	Формировать новые конкурентоспособные идеи в области теории и практики ГИС и ГИС технологий или системной инженерии программного обеспечения. Разрабатывать методы решения нестандартных задач и новые методы решения традиционных задач. Организовывать взаимодействие коллективов, принимать управленческие решения, находить компромисс между различными требованиями как при долгосрочном, так и при краткосрочным планировании.
<i>Универсальные (общекультурные) компетенции</i>	
P9	Использовать на практике умения и навыки в организации исследовательских, проектных работ и профессиональной эксплуатации современного оборудования и приборов, в управлении коллективом.
P10	Свободно пользоваться русским и иностранным языками как средством делового общения.
P11	Совершенствовать и развивать свой интеллектуальный и общекультурный уровень. Проявлять инициативу, в том числе в ситуациях риска, брать на себя всю полноту ответственности.
P12	Демонстрировать способность к самостоятельному обучению новым методам исследования, к изменению научного и научно-производственного профиля своей профессиональной деятельности, способность самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе в новых областях знаний, непосредственно не связанных со сферой деятельности, способность к педагогической деятельности.

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Школа Информационных технологий и робототехники
Направление подготовки 09.04.02 «Информационные системы и технологии»
Отделение школы (НОЦ) Информационных технологий

УТВЕРЖДАЮ:
Руководитель ООП
_____ Марков Н.Г.
(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

магистерской диссертации

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

Студенту:

Группа	ФИО
8ИМ6А	Гайнудинову Тимуру Сохибназаровичу

Тема работы:

Анализ индуктивного метода прогнозирования временных рядов и поиск мета-характеристик для улучшения его репрезентативности	
Утверждена приказом директора (дата, номер)	№ 2091/с от 26.03.2018 г.

Срок сдачи студентом выполненной работы:	
--	--

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p>Исходные данные к работе</p> <p><i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	<p>Объект исследования: программная платформа, реализующая индуктивный подход к прогнозированию временных рядов.</p> <p>Предметы исследования: мета-характеристики, позволяющие улучшить результаты работы программной платформы.</p>
---	---

<p>Перечень подлежащих исследованию, проектированию и разработке вопросов</p> <p><i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ul style="list-style-type: none"> • Обзор методов анализа временных рядов • Программная платформа • Поиск новых мета-характеристик • Разработка и внедрение найденных мета-характеристик • Перенос мета-характеристик на Python
--	---

<p>Перечень графического материала</p>	<p>презентация из 27 слайдов в формате .pptx, 21 таблица, 10 рисунков</p>
---	---

Консультанты по разделам выпускной квалификационной работы

Раздел	Консультант
<p>Финансовый менеджмент, ресурсоэффективность и ресурсосбережение</p>	<p>Старикова Е.В.</p>
<p>Социальная ответственность</p>	<p>Король И.С.</p>

Названия разделов, которые должны быть написаны на русском и иностранном языках:

<ul style="list-style-type: none"> • Предметная область
<ul style="list-style-type: none"> • Обзор программной платформы

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

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
<p>Доцент ОИТ</p>	<p>Буркатовская Ю.Б.</p>	<p>к.ф. – м.н., доцент</p>		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
<p>8ИМ6А</p>	<p>Гайнудинов Тимур Сохибназарович</p>		

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Школа Информационных технологий и робототехники
Направление подготовки 09.04.02 «Информационные системы и технологии»
Уровень образования магистратура
Отделение школы (НОЦ) Информационных технологий
Период выполнения осенний/весенний семестр 2017/2018 учебного года

Форма представления работы:

магистерская диссертация

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

**КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
выполнения выпускной квалификационной работы**

Срок сдачи студентом выполненной работы:

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
28.01.18	Основная часть	75
15.05.18	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	15
30.05.18	Социальная ответственность	10

Составил преподаватель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Буркатовская Ю.Б.	к.ф. – м.н., доцент		

СОГЛАСОВАНО:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ	Марков Н.Г.	д.т.н.		

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

Студенту:

Группа 8ИМ6А	ФИО Гайнудинов Тимур Сохибназарович
------------------------	---

Школа Уровень образования	ИШИТР Магистратура	Отделение Направление/ специальность	Информационных технологий 09.04.02 Информационные системы и технологии / Системная инженерия программного обеспечения
-------------------------------------	------------------------------	---	---

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

<i>1. Стоимость ресурсов научного исследования (НИ): материально – технических, энергетический, финансовых, информационных и человеческих</i>	Оклад руководителя принять в соответствии с научной степенью и занимаемой должностью, оклад студента принять равным окладу инженера ТПУ. Среднедневную заработную плату рассчитывать при 6-дневной рабочей недели для научного руководителя, и при 5-дневной рабочей недели для студента. Коэффициент отчислений на уплату на единый социальный налог 30,2%. Коэффициент НДС принять в размере 18 % от суммы затрат на разработку и прибыли.
<i>2. Нормы и нормативы расходования ресурсов</i>	
<i>3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

<i>1. Анализ конкурентных технических решений</i>	Проведение пред-проектного анализа: оценка потенциальных потребителей, SWOT-анализ, FAST-анализ, построение диаграммы Исикавы, анализ при помощи технологии Quad.
<i>2. Планирование управления научно-техническим проектом</i>	Планирование этапов работы, определение календарного графика и трудоемкости разработки, расчет бюджета.
<i>3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования</i>	Проведение сравнительного анализа характеристик вариантов исполнения проекта, вычисление эффективности разработки.

Перечень графического материала (с точным указанием обязательных чертежей):

1. Карта сегментирования рынка услуг по прогнозированию временных рядов
2. Оценочная карта для сравнения конкурентных технических решений
3. Классификация функций, выполняемых объектом исследования
4. Матрица смежности
5. Матрица количественных соотношений функций
6. Причинно-следственная диаграмма Исикавы
7. Оценочная карта для сравнения конкурентных технических решений
8. SWOT-анализ
9. Перечень работ
10. Трудозатраты на выполнение проекта
11. Линейный график работ
12. Нарастание технической готовности работы и удельный вес каждого этапа
13. Расчет затрат

14. Сравнительная оценка характеристик вариантов исполнения проекта
15. Сравнительная эффективность разработки

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

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН	Старикова Е.В.	к.ф.н		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ИМ6А	Гайнудинов Тимур Сохибназарович		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»**

Студенту:

Группа	ФИО
8ИМБА	Гайнудинов Тимур Сохибназарович

Школа	ИШИТР	Отделение школы	ИТ
Уровень образования	Магистратура	Направление/специальность	09.04.02 Информационные системы и технологии / Системная инженерия / программного обеспечения

Исходные данные к разделу «Социальная ответственность»:

<p>1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения</p>	<p>В выпускной квалификационной работе проводится анализ индуктивного метода прогнозирования временных рядов и проводится поиск новых мета-характеристик для улучшения его репрезентативности.</p>
---	--

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

<p>1. Профессиональная социальная безопасность. 1.1. Анализ вредных и опасных факторов, которые может создать объект исследования. 1.2. Анализ вредных и опасных факторов, которые могут возникнуть на рабочем месте при проведении исследований. 1.3. Обоснование мероприятий по защите исследователя от действия опасных и вредных факторов.</p>	<p>Анализ выявленных вредных факторов: -повышенный уровень электромагнитных излучений; - повышенная напряжённость электрического поля; Анализ выявленных опасных факторов, которые может создать объект исследования: - электрический ток (источником является ПК);</p>
<p>2. Экологическая безопасность. 2.1. Анализ влияния объекта исследования на окружающую среду. 2.2. Анализ влияния процесса исследования на окружающую среду. 2.3. Обоснование мероприятий по защите окружающей среды.</p>	<p>При выполнении работы влияние на атмосферу и гидросферу не происходит. Воздействие на литосферу – образование отходов при печати документов.</p>
<p>3. Безопасность в чрезвычайных ситуациях. 3.1. Анализ вероятных ЧС, которые может инициировать объект исследований. 3.2. Анализ вероятных ЧС, которые могут возникнуть на рабочем месте при проведении исследований. 3.3. Обоснование мероприятий по предотвращению ЧС и разработка порядка действия в случае возникновения ЧС.</p>	<p>В помещении возможно ЧС техногенного характера – пожар (возгорание).</p>

<p>4. Правовые и организационные вопросы обеспечения безопасности.</p> <p>4.1. Специальные (характерные для проектируемой рабочей зоны) правовые нормы трудового законодательства.</p> <p>4.2. Организационные мероприятия при компоновке рабочей зоны.</p>	<p>Требования к организации оборудования рабочих мест с ПК регулируется СанПиН 2.2.2/2.4.1340-03.</p> <p>Правовые вопросы регулируются трудовым кодексом РФ.</p>
--	--

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

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент отд. Кид	Король Ирина Степановна	Канд. хим. наук		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ИМ6А	Гайнудинов Тимур Сохибназарович		

Реферат

Магистерская диссертация состоит из 117 страниц, содержит 10 рисунков, 25 литературных источников.

Ключевые слова: анализ временного ряда, интеллектуальный анализ данных, мета-характеристики, МГУА, индуктивный метод прогнозирования, показатель Херста, экспонента Ляпунова, SSA.

Объектом исследования магистерской диссертации является программная платформа, предназначенная для анализа временных рядов.

Предметом исследования магистерской диссертации являются мета-характеристики, присутствующие и отсутствующие в данной программной платформе.

Цель работы – выявить репрезентативные мета-характеристики для увеличения точности программной платформы.

В процессе исследования проводились изучение и описание объекта исследования, методов и средств анализа временных рядов, а также изучение, тестирование и поиск мета-характеристик. Полученные результаты позволяют увеличить базу мета-характеристик временного ряда, а также улучшить качество прогнозирования программной платформы.

Определения, обозначения, сокращения, нормативные ссылки

МГУА – метод группового учета аргументов

Временной ряд – набор показателей процесса, зависящих от времени

ИАД – интеллектуальный анализ данных

Мета-характеристика – определенные неочевидные знания, имеющие практическую ценность

Оглавление

Введение	15
1. Предметная область.....	17
1.1. Временной ряд	17
1.2. Прогнозирование временных рядов	18
1.3. Интеллектуальный анализ данных	20
2. Обзор программной платформы	23
2.1. Общие сведения	23
2.2. База мета-характеристик.....	23
2.3. Постановка задачи добавления мета - характеристик	24
3. Метод группового учета аргументов	28
3.1. Описание.....	28
3.2. Применение в платформе	29
4. Chaos.....	31
4.1. Экспонента Ляпунова	31
4.2. Алгоритм.....	32
4.3. Допущения.....	33
4.4. Результаты тестов	33
5. Discontinuity	36
5.1. Коэффициент Херста.....	36
5.2. Алгоритм.....	37
5.3. Результаты тестов	37
6. Noise Influence	40
6.1. SSA.....	40
6.2. Алгоритм.....	41
6.3. Допущения.....	42
6.4. Результаты тестов	42
7. Питон	44
7.1. Перенос на Питон	44
7.2. Functional and Pythonic way	45
8. Шум-устойчивость	47
8.1. Шум-устойчивость модуля Discontinuity	47
8.2. Шум-устойчивость модуля Noise Influence	48
8.3. Шум-устойчивость модуля Chaos	49
9. Финансовый менеджмент, ресурс эффективность и ресурсосбережение	50
9.1. Потенциальные потребители результатов исследования.....	50

9.1.1.	Анализ конкурентных технических решений с позиции ресурс эффективности и ресурсосбережения	51
9.1.2.	FAST-анализ.....	54
9.1.3.	Диаграмма Исикава	56
9.1.4.	Технология Quad.....	57
9.1.5.	SWOT анализ	60
9.2.	Планирование управления научно-техническим проектом.....	62
9.2.1.	Структура работ в рамках научного исследования.....	62
9.2.2	Продолжительность этапов работ.....	63
9.3	Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	74
10	Социальная ответственность.....	78
10.1	Производственная безопасность	78
10.1.1	Повышенный уровень электромагнитных излучений; повышенная напряжённость электрического поля	79
10.1.2	Электрический ток (источник: ПК).....	81
10.2	Экологическая безопасность.....	82
10.2.1	Анализ воздействия на окружающую среду.....	82
10.3	Безопасность в чрезвычайных ситуациях	83
10.3.1	Пожарная безопасность	83
10.4	Правовые и организационные вопросы обеспечения безопасности	85
10.4.1	Организационные мероприятия обеспечения безопасности	85
10.4.2	Правовые мероприятия обеспечения безопасности.....	86
	Заключение	89
	Список использованных источников	90

Введение

Во многих областях жизни человека присутствуют процессы, которые возможно наблюдать и отслеживать во времени. Одновременно с этим возникает задача прогнозирования поведения процесса. Однако, при существовании большого количества методов прогнозирования временных рядов, хорошо проявляющих себя на одних типах процессов и плохо прогнозирующих на других, отсутствуют средства автоматического определения наилучшего алгоритма для прогнозирования определенного процесса. На данный момент эта задача лежит на экспертах, что является сильно затратным и не автоматизированным решением.

Орловым Андреем Александровичем, сотрудником компании Econophysics, была разработана программная платформа, предназначенная для исследования и прогнозирования временных рядов, связанных с финансами.

Econophysics - международная компания, которая оказывает услуги крупнейшим мировым игрокам сферы финансовых и банковских услуг.

Область специализации компании - разработка и валидация сложных математических моделей, алгоритмов и программных продуктов для финансовой сферы, поддерживающих комплексные финансовые инструменты и предоставляющих глубокие возможности для анализа состояния портфеля и оценки рисков.

Программная платформа представляет собой комплекс моделей, алгоритмов и программ интеллектуального анализа данных, реализующий подход мета-обучения и основанный на методе группового учета аргументов.

В ходе тестирования программной платформы ее автором было выявлено, что при вычислении мета-характеристик временных рядов частично теряется информация об периоде и шуме исследуемого временного ряда. Исходя из этого и появляется задача поиска новых мета-характеристик для улучшения репрезентативности программной платформы.

В данной работе приводится обзор таких понятий, как временной ряд, интеллектуальный анализ данных и метод группового учета аргументов, дается краткий обзор вышеупомянутой программной платформы, демонстрируются мета-характеристики, уже реализованные в платформе, а также приведены мета-характеристики, выбранные для расширения платформы.

Также в отчете приведены результаты проверки модулей на устойчивость к шуму и перенос модулей на язык программирования Python.

1. Предметная область

1.1. Временной ряд

Временной ряд представляет собой набор статистических данных, отображающих исследуемый процесс и привязанных ко времени. Каждый элемент этих данных называют измерением, либо уровнем относительно определенного момента времени. Исходя из этого, порядок элементов во временном ряду имеет важное значение. В этом и заключается существенное отличие временного ряда от простой выборки данных.

Так как временные ряды представляют собой набор двух элементов, а именно:

- числовое значение;
- временная отметка, которой соответствует числовое значение

временные ряды бывают различных видов:

- одномерные и многомерные временные ряды, отличающиеся количеством показателей на один момент времени;
- моментные и интервальные временные ряды, отличающиеся числовым значением на определенный момент времени. В моментных временных рядах числовое значение характеризует процесс на определенный момент времени, а в интервальных временных рядах – за определенный период времени;
- равноотстоящие и неравноотстоящие временные ряды, отличающиеся типом интервалов между временными отметками; интервалы могут быть равными или отличными друг от друга;
- полные и неполные временные ряды, т.е., с отсутствием или наличием пропущенных значений .

Анализ временных рядов представляет собой совокупность различных статистических и математических методов, применяемых для выделения статистических характеристик, структуры временного ряда и некоторых иных

показателей. Статистические характеристики дают представление о свойствах временного ряда. Понятие структуры временного ряда важно для построения математической модели исследуемого процесса, являющегося источником временного ряда. В задачи анализа временного ряда также входит задача выделения компонент временного ряда, таких как трендовая, циклическая, сезонная составляющие и шум, который представляет собой остаточную информацию, не поддающуюся объяснению.

Все вышеперечисленные данные используются в задачах прогнозирования временных рядов, возникающих при проблеме эффективного принятия решений. Прогноз представляет собой построение модели временного ряда, основываясь на истории исследуемого процесса для получения вероятного значения будущих данных. Задача прогнозирования временных рядов является одной из важнейших задач при работе с временными рядами. Данная задача возникает почти во всех областях жизни человека, например, прогнозирование погоды, прогнозирование стоимости ценных бумаг, и т.д. [1]

1.2. Прогнозирование временных рядов

На сегодняшний день существует огромное количество моделей прогнозирования временных рядов. Среди них есть как простые (к примеру, линейная регрессия), так и сложные (модель на основе передаточных функций) [2]. Их можно классифицировать на несколько групп:

- регрессионные модели – модели, основанные на поиске зависимости между одной переменной и множеством других факторов
 - простая линейная регрессия
 - множественная регрессия
 - нелинейная регрессия

- авторегрессионные модели - модель, основанные на предположении, что следующие состояния процесса зависят от нескольких предыдущих состояний системы
 - ARIMAX - autoregression integrated moving average extended
 - GARCH - generalized autoregressive conditional heteroskedasticity
 - ARDLM - autoregression distributed lag model
- модели экспоненциального сглаживания – модели, основанные на сглаживании временного ряда с помощью взвешенной скользящей средней
 - экспоненциальное сглаживание
 - модель Хольта
 - модель Хольта - Винтерса
- модель по выборке максимального подобия – модель, основанная на идее повторяемости процесса, но с измененными свойствами
- модель на нейронных сетях – модель, представляющая собой нейронную сеть
- модель на цепях Маркова – модель, основанная на предположении, что следующее состояние процесса зависит только от текущего и никак не связано с предыдущими
- модели на классификационно-регрессионных деревьях -модель, основанная на построении деревьев классификации, регрессионных деревьев либо классификационно-регрессионных деревьев
- модель на основе генетического алгоритма - модель, основанная на построении генетического алгоритма
- модель на опорных векторах – модель, основанная на идее переноса временного ряда в пространство, размерность которого

превосходит исходное, и поиска разделяющей гиперплоскости с максимальным зазором в этом пространстве

1.3. Интеллектуальный анализ данных

В современном мире объемы собираемых и анализируемых данных просто огромны. Человек не в состоянии проанализировать такое количество данных самостоятельно, но задача анализа этих данных требует решения, ведь в этих данных скрыты полезные знания, закономерности, зависимости, которые могут быть использованы при принятии решений. Автоматический анализ данных называют интеллектуальным анализом данных, или Data mining [3].

Data mining (дословно «добыча данных») – процесс обнаружения нетривиальной полезной информации в больших объемах данных, не предъявляемой в явном виде, но полезной в практическом смысле. Полученные закономерности и тенденции невозможно обнаружить при традиционном просмотре данных в силу сложности связей и чрезмерного объема данных.

Основные инструменты, применяемые в интеллектуальном анализе данных для решения задач:

Обучение с учителем – данный метод подразумевает собой получение на входе пары значений: какой-либо объект и соответствующий ему ответ. Задачей ставится поиск и выведение функциональной зависимости между объектом и ответом. Задачи, решаемые с помощью обучения с учителем, делятся на:

- классификация – разделение заданного набора данных на группы по заранее определенным группам
- регрессия – выявление закономерностей между определенными признаками или данными
- ранжирование – получение результатов сразу для нескольких объектов с последующей сортировкой

- прогнозирование – получение будущих состояний исследуемого объекта

Обучение без учителя – данный метод на основе полученных объектов, в условиях отсутствия соответствующих этим объектам ответов, должен установить связь между объектами. Задачи, решаемые с помощью обучения без учителя, делятся на:

- кластеризация – группировка объектов на основе их признаков, свойств, параметров. Важным условием является то, что объекты внутри кластеров должны быть похожи, и должны отличаться от объектов из других кластеров
- ассоциация – выявление закономерностей между данными
- фильтрация выбросов – обнаружение нетипичных объектов в общей выборке
- построение доверительной области – выделение области, содержащей нужную долю выборки, минимизируя объем области
- сокращение размерности – переход из одного пространства признаков в другое, меньшее по размерности, при этом теряя как можно меньше информации
- заполнение пропущенных значений – замена недостающих данных при помощи построения прогноза

Частичное обучение – промежуточное звено между обучением с учителем и обучением без учителя.

Трансдуктивное обучение – метод, решающий задачи прогнозирования данных на одной частной выборке для другой частной выборки.

Обучение с подкреплением – метод, оперирующий с данными типа «ситуация» и «принятое решение». Ответом для метода является функция, определяющая «правильность» принятия решения.

Динамическое обучение – разница от других методов обучения состоит в том, что данные поступают потоком, что вынуждает одновременно принимать решение и дообучать модель на новых данных.

Активное обучение - отличается тем, что обучаемый имеет возможность самостоятельно назначать следующий прецедент, который станет известен.

Мета – обучение – метод, в котором прецедентом являются уже решенные задачи обучения. Задачи, решаемые при помощи мета-обучения, делятся на:

- многозадачное обучение – набор одновременно решаемых задач обучения
- индуктивный перенос - опыт решения отдельных частных задач обучения по прецедентам переносится на решение последующих частных задач обучения

2. Обзор программной платформы

2.1. Общие сведения

Исследуемая программная платформа была полностью разработана и реализована кандидатом технических наук Орловым Андреем Александровичем. В перспективе платформа представляет собой не только научный интерес, как автоматический индуктивный метод прогнозирования, так и промышленный, к примеру, для автоматического прогнозирования данных в различных областях либо как альтернативное экспертному мнению решение.

Данная платформа является программным комплексом для реализации системы мета-обучения с использованием метода группового учета аргументов. Платформа представляет собой четырехуровневую систему. На последнем уровне система строит прогноз на основе параметров модели и коэффициентов, полученных на уровне выше и добавляет результаты в базу данных. На третьем уровне на основе наилучшего алгоритма решения и параметров модели вычисляются коэффициенты для выбранной модели. На втором уровне выделяются мета-характеристики временного ряда и на основе информации из базы данных на уровне выше выбирается наилучший алгоритм и параметры для прогнозирования анализируемого временного ряда. На первом уровне находится база данных, в которой содержатся результаты для каждого алгоритма, со всеми возможными, до заданного ограничения, параметрами, а также ошибка прогнозирования для каждого метода. В результате система выдает 8 наилучших алгоритмов для прогнозирования анализируемого временного ряда, оценивая по критерию эффективности. Подробное описание можно найти по ссылке [4].

2.2. База мета-характеристик

На данный момент в системе представлено 8 мета-характеристик временного ряда:

1. **Length** – Длина - длина временной последовательности (число точек).
2. **Noise** – Шум - средняя величина абсолютного изменения за шаг.

$$F_{\text{noise}} = \frac{1}{T} \sum_{t=1}^{T-1} |x_{t-1} - x_t|$$

3. **Trend** – Тренд - коэффициент наклона прямой линейного тренда.

$$F_{\text{trend}} = a, x_t = a * t + b$$

4. **Period** – Период - период гармонического сигнала, рассчитываемый согласно критерию баланса ординат из предположения наличия единственной гармоники во входной временной последовательности.

Осуществлялось построение прогнозирующей модели с использованием комбинаторного алгоритма метода МГУА с ограничением на 1-ю (land1, линейная модель) и 7-ю (land2, нелинейная модель) степень полинома. Обучающая выборка составляла первую половину точек временной последовательности, проверочная – вторую. В качестве внешнего критерия использовался критерий регулярности CR.

5. **Land1in** - Значение CR на обучающей выборке данных
6. **Land1out** - Значение CR на проверочной выборке данных
7. **Land2in** - Значение CR на обучающей выборке данных
8. **Land2out** - Значение CR на проверочной выборке данных

2.3. Постановка задачи добавления мета - характеристик

Задача поиска новых мета-характеристик возникла в ходе тестирования существующего программного обеспечения. В результате тестов выяснилось, что при подсчете мета-характеристик для различных временных рядов, теряется важная информация о некоторых свойствах временного ряда, а именно, о шуме и периоде.

Исходя из этого, задача улучшения адекватности результатов метода сводилась к анализу проблемы потери информации во время подсчета мета-характеристик и поиску таких мета-характеристик, которые помогут лучше отслеживать метазнания о шумовой и периодической составляющих анализируемого временного ряда.

В ходе поиска был выявлен ряд характеристик [5], предложенных в различных источниках и статьях:

- Number of instances
- Number of features
- Number of classes
- Dataset dimensionality
- Standard deviation
- Variation coefficient
- Correlation coefficient
- Skewness
- Kurtosis
- Average normalized features entropy
- Normalized class attribute entropy
- Maximal mutual information between the attribute and the class
- Average mutual information between the attribute and the class
- Ratio of signal noise
- Number of equivalent features
- Weight sum on the full dataset
- Minimum weight sum
- Average weight sum
- Standard deviation of weight sum
- Number of neighbors for the full dataset
- Minimum number of neighbors
- Maximum number of neighbors

- Average number of neighbors
- Standard deviation of the number of neighbors

Критериями выбора новых мета-характеристик являлись:

- данные, полученные в результате вычисления мета-характеристики, должны быть нетривиальными и ранее неизвестными
- данные, полученные в результате вычисления мета-характеристики, должны содержать в себе информацию о шумовой либо периодичной составляющей анализируемого процесса

Перечисленные мета-характеристики, в силу несоответствия критериям, выставленным для поиска новых мета – характеристик, были отвергнуты.

По результатам проведенного анализа были предложены следующие мета-характеристики:

Chaos – Хаос – хаотическая составляющая, вычисляемая при помощи экспоненты Ляпунова.

Динамический хаос — явление в теории динамических систем, при котором поведение нелинейной системы выглядит случайным несмотря на то, что оно определяется детерминистическими законами.

Noise Influence – Влияние шума – данная характеристика вычисляется при помощи спектрального сингулярного анализа (SSA).

Обнаружение влияния шума проводится с помощью вычисления разницы пространства шума от оцененного подпространства сигнала. Соответственно, если разница начинает расти, это свидетельствует об сильной зашумленности.

Discontinuity – Скачкообразность – определяется показателем Херста.

Одной из наиболее важных задач мониторинга технического объекта является задача своевременного обнаружения существенных изменений

физических величин (переменных состояния), характеризующих состояние объекта. Скачкообразное изменение даже одной переменной, как правило, сигнализирует о возможности серьезной трансформации состояния объекта, которое может привести к нештатной ситуации или аварии.

3. Метод группового учета аргументов

3.1. Описание

Метод группового учета аргументов (МГУА) представляет собой метод массовой селекции регрессионных моделей до тех пор, пока не будет найдена модель оптимальной сложности. Сложность модели, в понимании метода, является собой количество параметров модели. Для получения модели используются исходные элементы. Для определения оптимальности модели используются различные внешние критерии, позволяющие отслеживать качество получаемых моделей.

Один из авторов этого метода А.Г. Ивахненко пишет: «Осуществляется целенаправленный перебор многих моделей-претендентов различной сложности по ряду критериев. В результате находится модель оптимальной структуры в виде одного уравнения или системы уравнений. Минимум критерия селекции определяет модель оптимальной структуры» [6].

МГУА представляет собой индуктивный подход к поиску оптимальной модели. Перебор моделей осуществляется от минимально простых и с каждым проходом цикла алгоритма метод увеличивает сложность модели, пока показатель внешнего критерия для моделей не начнет ухудшаться.

Данный метод позволяет найти оптимальную сложность модели, не прибегая к «экспертной» помощи, основываясь не на абстрактном опыте эксперта, а на обрабатываемых данных. Задача поиска оптимальной сложности модели очень значима, так как применив слишком простую модель к данным, имеющим сложные зависимости, есть риск потерять большой объем информации. Аналогична и обратная ситуация, когда описываемый простой процесс вписывается в сложную модель, порождая несуществующие зависимости и лишнюю информацию, дающие совершенно отдаленные от реальности результаты при дальнейшем использовании модели.

Существует много видов алгоритмов МГУА, но базовая концепция представляет собой алгоритм из 5 шагов:

Шаг 1. Выбираются опорные функции – то есть общий вид моделей, которые будут перебираться в ходе работы алгоритма. Обычно прибегают к полиному Колмогорова – Габора:

$$Y(x_1, \dots, x_n) = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ijk} x_i x_j x_k + \dots$$

Шаг 2. Строятся все вариации модели для всех или нескольких аргументов с применением выбранной опорной функции. Для каждой модели определяются ее коэффициенты методом регрессионного анализа.

Шаг 3. Происходит селекция моделей, то есть выбираются несколько лучших моделей.

Шаг 4. Проверяется, достигнуто ли ограничение на максимальную сложность модели и оценивается насколько модель «хороша» и лучше предыдущих. Если достигаются вышеперечисленные условия, алгоритм прекращает работу.

Шаг 5. Если алгоритм не закончил работу, модели, полученные на шаге 3, используются как аргументы для построения новых, более сложных, моделей, и повторяются шаги алгоритма, начиная с шага 2.

3.2. Применение в платформе

Метод группового учета аргументов был рассмотрен, так как он очень широко применяется в исследуемой платформе.

В первую очередь, МГУА в платформе применяется для построения мета-модели исследуемого временного ряда. Это происходит на втором уровне работы платформы, на котором исходя из насчитанных мета-данных насчитывается мета-модель.

Во-вторых, различные вариации алгоритма МГУА представлены в наборе алгоритмов, среди которых производится отбор лучшего алгоритма, и

в последствии, строится прогноз при помощи этого алгоритма. Среди этих алгоритмов представлены:

- комбинаторный
- полиномиальный
- гармонический
- дважды полиномиальная сеть

В-третьих, МГУА используется для вычисления мета-характеристик:

- Land1in
- Land1out
- Land2in
- Land2out

4. Chaos

4.1. Экспонента Ляпунова

Экспонента Ляпунова характеризует экспоненциальную скорость расхождения близких траекторий во времени. При определении экспоненты Ляпунова большое отклонение от нуля свидетельствует о том, что исследуемый процесс чувствителен к значениям начальных условий, то есть близкие траектории со временем отдаляются друг от друга. Таким образом экспонента Ляпунова позволяет сделать вывод о характере динамической системы с точки зрения присутствия в ней хаотической составляющей [7].

Общая форма уравнения, описывающего эволюцию динамической системы (1):

$$\frac{dx}{dt} = F(x)$$

где x – n -мерный вектор, описывающий конфигурацию системы. Пусть $x_1(t)$ и $x_2(t)$ – два линейно независимых решения эволюционного уравнения, такие что $x_1(t_0) \approx x_2(t_0)$. Обозначим $\beta(t) = x_2(t) - x_1(t)$, тогда

$$F(x_2(t)) = F(x_1(t) + \beta(t)) \approx F(x_1(t)) + \beta(t) \frac{\partial F_i}{\partial x_j}$$

Принимая во внимание эволюционное уравнение, заменяем все вхождения функции $F(x)$ на dx/dt :

$$\frac{dx_2}{dt} - \frac{dx_1}{dt} = \frac{d\beta}{dt} = \beta(t) \frac{\partial F_i}{\partial x_j}$$

Таким образом, внутри некоторой n -мерной сферы линеаризация вокруг точки фазового пространства $x_1(t_0)$ получаем систему обыкновенных линейных дифференциальных уравнений:

$$\frac{d\beta_i(t)}{dt} = A_{ij}\beta_j(t)$$

Если все собственные числа матрицы A_{ij} различны, то общее решение системы имеет вид:

$$\beta(t) = C_{ij} \begin{pmatrix} e^{\lambda_1 t} \\ e^{\lambda_2 t} \\ \dots \\ e^{\lambda_n t} \end{pmatrix}$$

где коэффициенты матрицы C_{ij} имеют связи друг с другом [8].

4.2. Алгоритм

Вычисление мета-характеристики Chaos основано на результатах вычисления экспоненты Ляпунова.

Мета-характеристика Chaos вычисляется следующим способом:

- На первом шаге алгоритм обрабатывает временной ряд и вычисляет обобщенные переменные.
 - Под обобщенными переменными понимаются первая, вторая, третья и четвертая производная.
- Далее в алгоритме применяется преобразование Фурье на текущий временной ряд.
 - Преобразование Фурье применяется для вычисления периода временного ряда
- На следующем шаге вычисляется среднее количество соседних точек, попадающих в фазовый шар, представляющий собой область, где линеаризация уравнения эволюции системы привносит приемлемую ошибку.
 - Если среднее количество точек меньше 1, то выставляется значение равное 1.
- На предпоследнем шаге в алгоритме производится поиск ближайших соседей.

- При чем точка начала поиска соседа задается с исследуемой точки + период, вычисленный на третьем шаге
- И на последнем шаге при помощи полученных данных вычисляется экспонента Ляпунова.

$$\lambda \cong \frac{1}{\Delta t} \ln \frac{\varepsilon(t)}{\varepsilon(t_0)}$$

4.3. Допущения

Наблюдаемый процесс и его производные образуют фазовое пространство.

Поскольку делается явное априорное допущение о размерности фазового пространства (5 вариантов одного и того же допущения – размерность фазового пространства от 1 до 5), то на каждое такое допущение вычисляется собственный показатель Ляпунова.

4.4. Результаты тестов

Алгоритмы было решено проверить на исторических данных котировок, взятых с сайта finam.ru.

На данных котировок отношения евро к доллару алгоритм дал результат в 1.6966, что означает присутствие слабой чувствительности к значениям начальных условий. (рис 1)

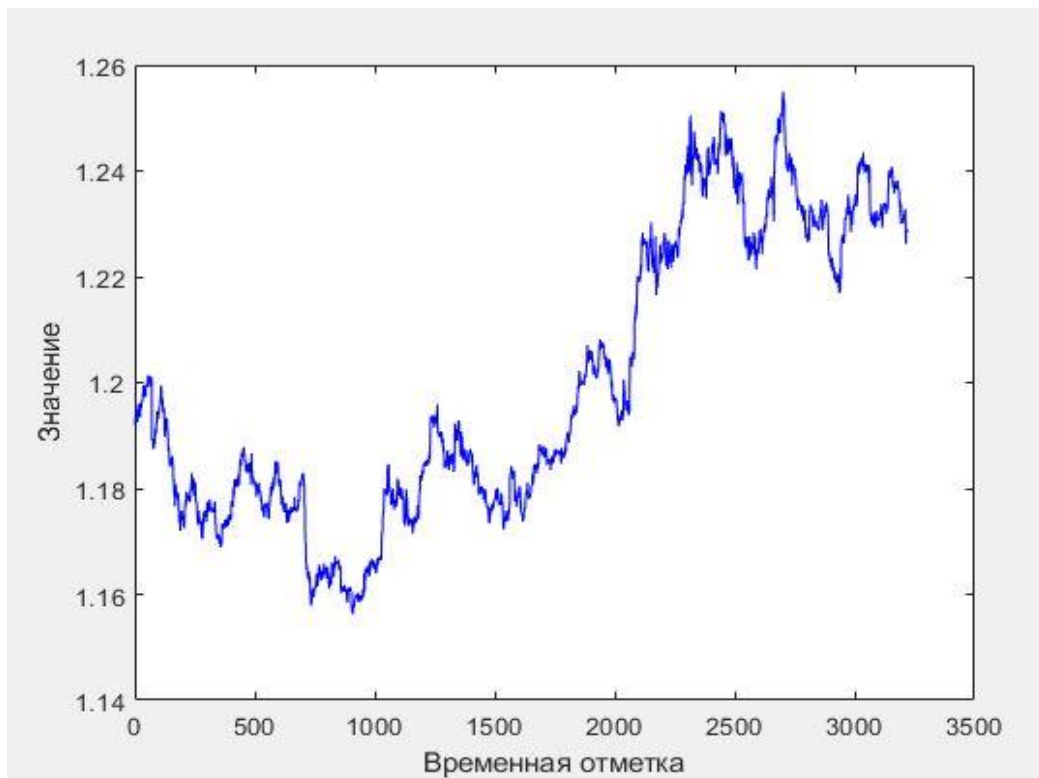


Рис 1. Исторические данные котировок EUR/USD

На данных котировок цены на акции Сбербанка алгоритм дал результат в 2.1422, что означает присутствие слабой чувствительности к значениям начальных условий. (рис 2)

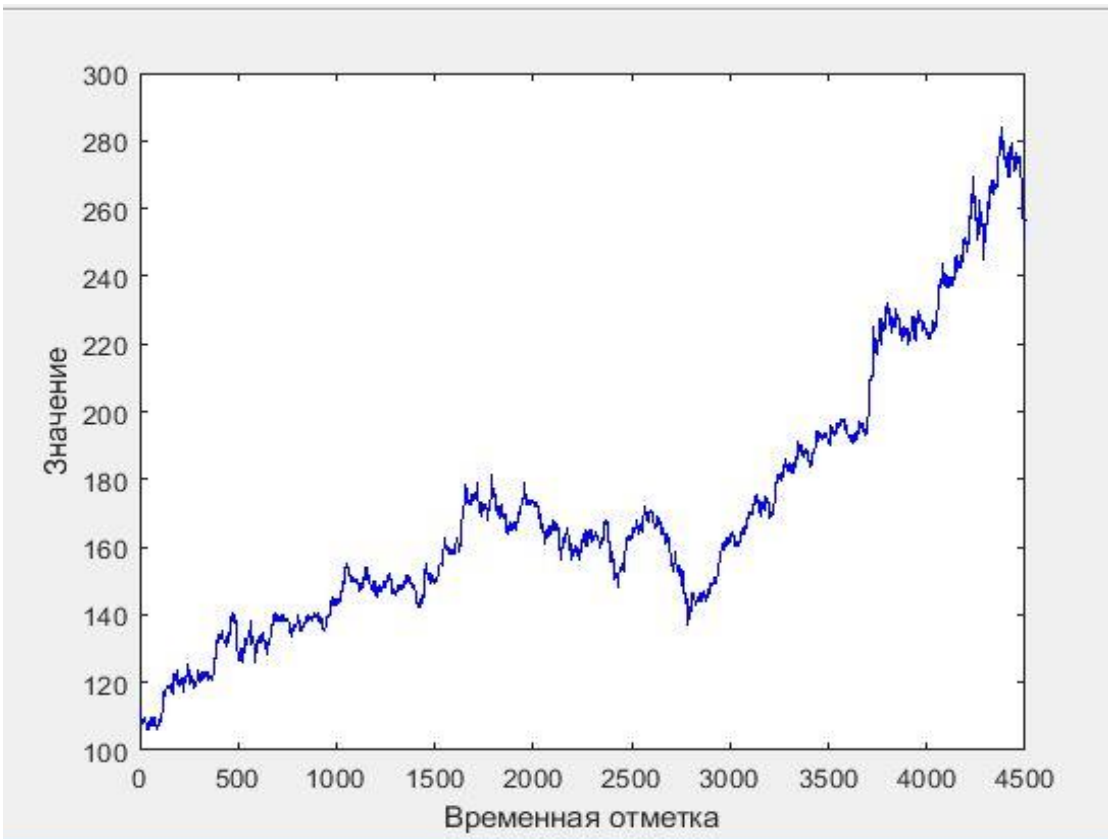


Рис 2. Исторические данные котировок цены акций Сбербанка

5. Discontinuity

5.1. Коэффициент Херста

За последние несколько лет в анализе временных рядов стал очень популярен так называемый метод нормированного размаха. При помощи этого метода ведется анализ процессов в различных областях [9].

Вычисление нормированного размаха происходит следующим способом. Пусть $x(t)$ – ряд значений, полученных в дискретные моменты времени t в течении промежутка времени τ , $mean(x(\tau))$ – среднее значение.

$$S(\tau) = \sqrt{\frac{1}{\tau} \sum_{i=1}^{\tau} [x(t_i) - mean(x(\tau))]^2} - \text{стандартное отклонение } x.$$

Следующим шагом выводится накопившееся отклонение случайной величины $x(t)$ от ее среднего значения для каждого момента t :

$$X(t, \tau) = \sum_{u=1}^t [x(u) - mean(x(\tau))]$$

Подсчитывается размах значений, получаемый при помощи разности максимального и минимального значений $X(t, \tau)$.

$$R = \max X(t, \tau) - \min X(t, \tau)$$

Отношение R/S позволяет обезразмерить значение и дает возможность сравнивать размах для различных случайных величин. При первых опытах работы с размахом было выявлено, что его можно описать отношением:

$$\frac{R}{S} = \left(\frac{\tau}{2}\right)^H$$

где H – это и есть показатель Херста.

Показатель Херста – показатель, используемый для определения характера изменения тенденции временного ряда. Последовательности считаются сохраняющими тенденцию, или персистентными, если показатель Херста для них более 0,5. Это означает, что убывание в прошлом скорее всего приведет к убыванию в дальнейшем, и наоборот. Аналогично,

последовательности считаются антиперсистентными, если показатель Херста для них менее 0,5. Это означает, что любая тенденция стремится смениться на противоположную. При значении показателя Херста равном 0,5 считается что тенденция не выражена явно [10].

Для природных процессов, к примеру, накопление водных ресурсов, статистически было выявлено что значения показателя Херста варьируются вблизи значений 0,72-0,73.

5.2. Алгоритм

Вычисление мета-характеристики Discontinuity основано на результатах вычисления показателя Херста.

Мета-характеристика Discontinuity вычисляется следующим способом:

- На первом шаге алгоритм вычисляет стандартное отклонение.
- Далее алгоритм вычисляет размах значений ряда.
- На следующем шаге при помощи формулы вычисляется показателя Хёрста.

5.3. Результаты тестов

Алгоритмы было решено проверить на исторических данных котировок, взятых с сайта finam.ru.

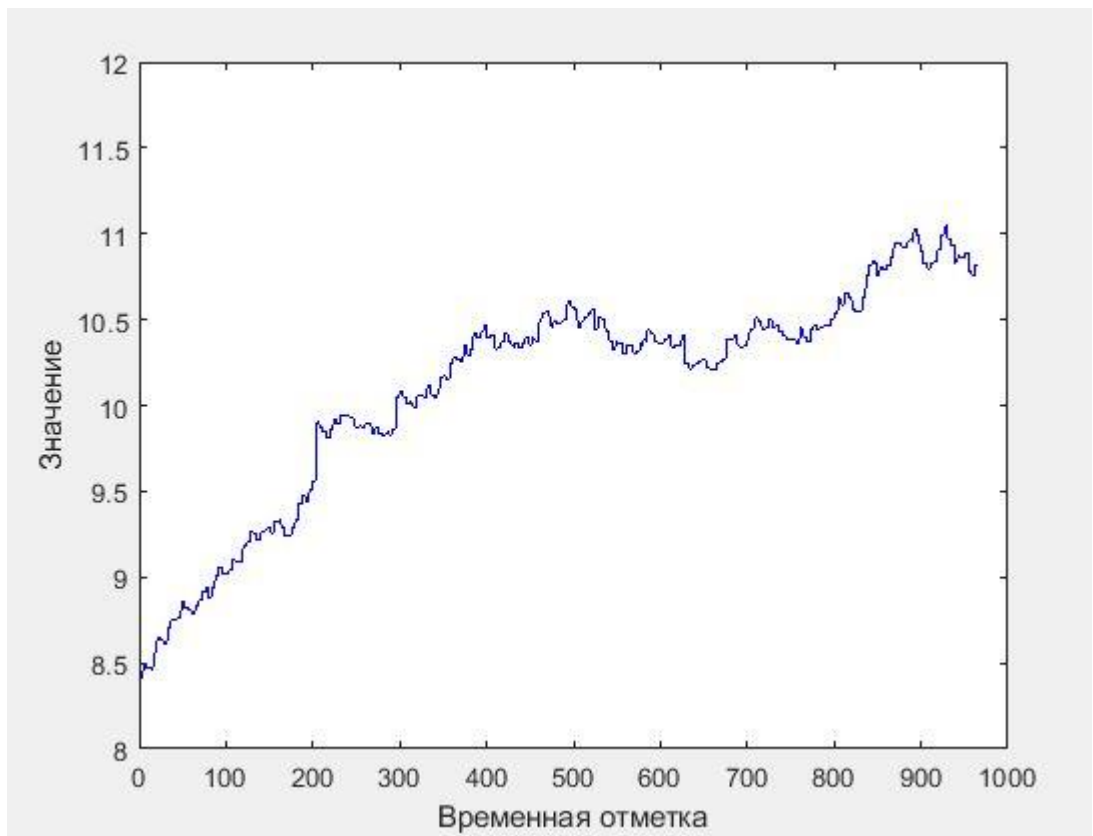


Рис 3. Исторические данные котировок EUR/TJS

На данных котировок отношения евро к сомони алгоритм дал результат в 0.6021, что означает сохранение тенденции на временном ряду. (рис 3)

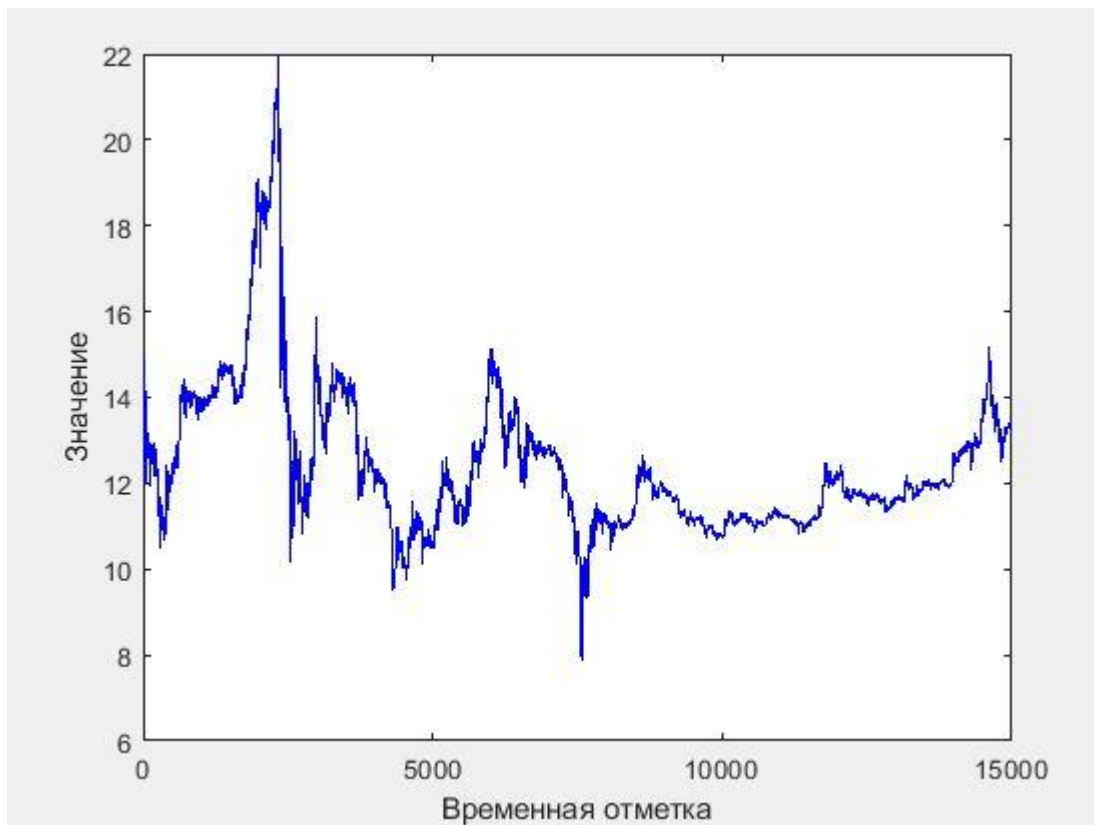


Рис 4. Исторические данные котировок ETH/ EUR

На данных котировок отношения евро к эфириуму алгоритм дал результат в 0.5420, что означает что какая-либо тенденция явно не выражена.
(рис 4)

6. Noise Influence

6.1. SSA

Спектральный сингулярный анализ — это метод, свободный от модели, применяемый в анализе структуры временных рядов, совмещает в себе достоинства многих других методов, в частности, анализа Фурье и регрессионного анализа. В то же время метод выделяется наглядностью в управлении.

Базовый вариант метода состоит в преобразовании одномерного ряда в многомерный с помощью однопараметрической сдвиговой процедуры (отсюда и название "Гусеница"), исследовании полученной многомерной траектории с помощью анализа главных компонент (сингулярного разложения) и восстановлении (аппроксимации) ряда по выбранным главным компонентам.

Таким образом, результатом применения метода является разложение временного ряда на простые компоненты: медленные тренды, сезонные и другие периодические или колебательные составляющие, а также шумовые компоненты[11].

Полученное разложение может служить основой прогнозирования как самого ряда, так и его отдельных составляющих. "Гусеница" допускает естественное обобщение на многомерные временные ряды, а также на случай анализа изображений.

Метод независимо разрабатывался в России, Великобритании и США (там его аналог получил название SSA - Singular Spectrum Analysis) и показал себя мощным средством исследования временных рядов (в основном, в метеорологии, гидрологии, климатологии, геофизике). [12]. Алгоритм анализа состоит из 4 этапов:

Этап 1. Вложение

На этом этапе из исходного временного ряда строится траекторная матрица с шириной окна L . Матрица состоит из $K = N - L + 1$ векторов вложения.

Этап 2. Сингулярное разложение.

На данном этапе вычисляется матрица $S = XX^T$, а затем вычисляются собственные тройки компонент.

Этап 3. Группировка

Полученные компоненты группируются на основе разложения.

Этап 4. Диагональное усреднение

Проводится диагональное усреднение по формуле (рис 5):

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m,k-m+2}^* & \text{для } 0 \leq k < L^* - 1, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+2}^* & \text{для } L^* - 1 \leq k < K^*, \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m,k-m+2}^* & \text{для } K^* \leq k < N. \end{cases}$$

Рис 5. Формула диагонального усреднения

6.2. Алгоритм

Вычисление мета-характеристики Noise Influence основано на результатах проведения SSA и суммирования расстояния между пространством сигнала и шума.

Мета-характеристика Noise Influence вычисляется следующим способом:

- На первом шаге временной ряд переводится в последовательность многомерных векторов.

- В результате получаем матрицу с шириной окна равной половине длины ряда
- Далее проводится сингулярное разложение
 - В результате получаем собственные тройки сингулярного разложения
- На следующем шаге проводится диагональное усреднение матрицы
 - Далее выделяется пространство сигнала
- На последнем шаге вычисляется значение зашумленности ряда
 - Значение вычисляется как разница между выделенным пространством сигнала и пространством шума

6.3. Допущения

Так как вопрос разделения пространства сигнала на составляющие компоненты нас не волнует, процесс группировки компонент был опущен в алгоритме.

6.4. Результаты тестов

На данных котировок отношения евро к доллару алгоритм дал результат ≈ 1401 , что означает сильную зашумленность данных. (рис 6)

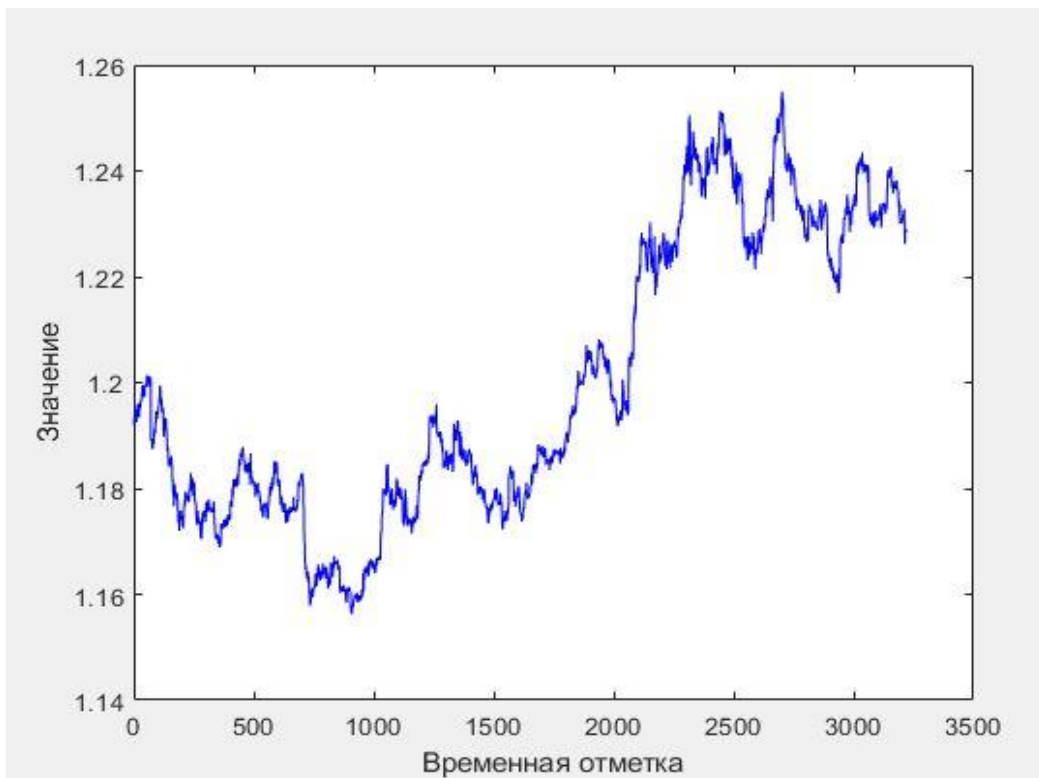


Рис 6. Исторические данные котировок EUR/USD

На специально сгенерированных данных алгоритм дал результат в ≈ 10 , что означает практическое отсутствие шума в данных. (рис 7)

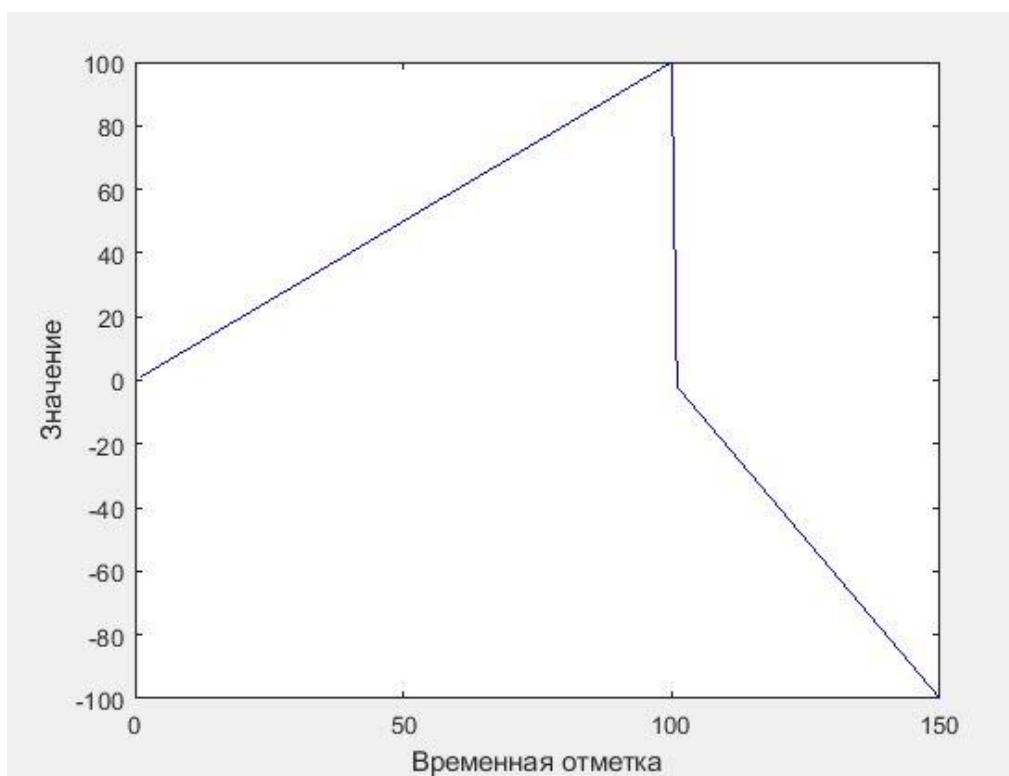


Рис 7. Проверочные данные

7. Питон

7.1. Перенос на Питон

Так как планировалось в дальнейшем перенести систему на язык программирования Python передо мной стояла задача переноса модулей на Python. Мною был проведен поиск и анализ существующих решений.

В ходе исследования был проведен поиск и рассмотрены альтернативы и готовые решения. Результаты приведены в таблице (таблица 1.):

Мета-характеристика	Метод	Matlab	Python
Discontinuity	Коэффициент Херста	+ [13]	-
Noise Influence	SSA	Не удовлетворяет задаче	Не удовлетворяет задаче
Chaos	Экспонента Ляпунова	Код непрозрачен, в комментариях много негативных отзывов. [14]	Решение на стадии разработки, на текущий момент v 0.4.1. [15]

Таблица 1. Результат анализа готовых решений

В ходе анализа готовых решений для мета-характеристики Discontinuity был найден подключаемый модуль на языке Matlab, на языке Python реализаций найдено не было.

В ходе анализа готовых решений для мета-характеристики Noise Influence были найдены решения на языках Matlab и Python, но в силу специфичности задачи они были отвергнуты.

В ходе анализа готовых решений для мета-характеристики Chaos была найдена библиотека, находящаяся только на стадии разработки. Так же была найдена реализация на Matlab, но в силу непрозрачности кода и негативных отзывов реализация была отвергнута.

7.2. Functional and Pythonic way

При переносе модулей на язык Python были соблюдены стандарты PEP8. Это позволило сохранить удобочитаемость кода и соответствовать Pythonic стилю написания кода [16]. Данный стиль описывается как «Дзен Питона» и звучит так:

Beautiful is better than ugly.

Explicit is better than implicit.

Simple is better than complex.

Complex is better than complicated.

Flat is better than nested.

Sparse is better than dense.

Readability counts.

Special cases aren't special enough to break the rules.

Although practicality beats purity.

Errors should never pass silently.

Unless explicitly silenced.

In the face of ambiguity, refuse the temptation to guess.

There should be one-- and preferably only one --obvious way to do it.

Although that way may not be obvious at first unless you're Dutch.

Now is better than never.

Although never is often better than **right** now.

If the implementation is hard to explain, it's a bad idea.

If the implementation is easy to explain, it may be a good idea.

Namespaces are one honking great idea -- let's do more of those!

Но, так как эти модули являются ничем иным, как математическими методами, возник вопрос перестройки модулей, используя функциональную парадигму программирования. Причиной такого решения было то, что характеристики функционального программирования приводят к более быстрой разработке более короткого и безошибочного кода [17].

При программировании, используя функциональную парадигму программирования, нужно акцентироваться на том, что должно быть вычислено, а не как.

Более того, высокие теоретики от компьютерной науки, логики и математики находят, что процесс доказательства формальных свойств для функциональных языков и программ много проще, чем для императивных.

8. Шум-устойчивость

Так как модули предназначены для работы с временными рядами, которые довольно часто бывают зашумленными – возникает вопрос о шум-устойчивости модулей.

В связи с этим были проведены тесты модулей для того, чтобы выявить насколько сильно будут отклоняться результаты модулей при добавлении шума в данные.

8.1. Шум-устойчивость модуля **Discontinuity**

Для тестирования мета-характеристики **Discontinuity** был использован временной ряд длиной в 3221 значение и с среднеквадратичным отклонением, равным 0.0270. Результатом вычисления на данном временном ряду является значение мета-характеристики, равное 0.5953.

Тестирование представляло собой 100 итераций с добавлением шума, представляющего собой случайные величины, распределенные по равномерному закону в интервале (0, 1) (рис 8).

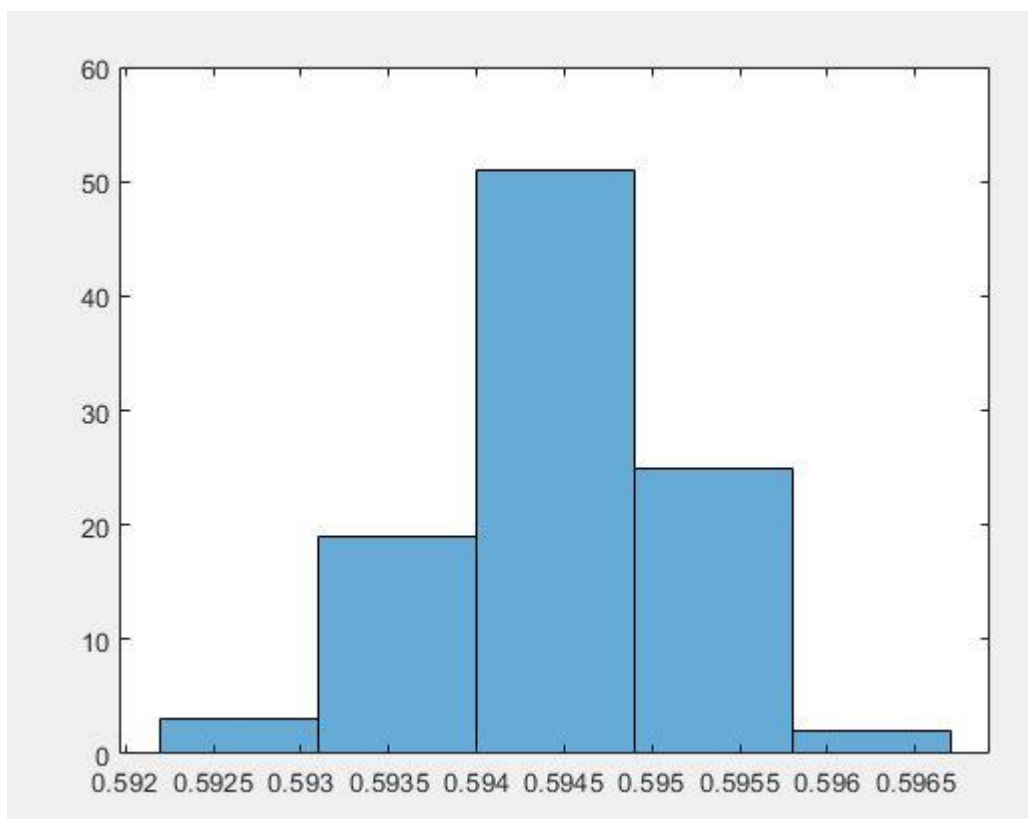


Рис 8. Гистограмма разброса значений после добавления шума

После добавления шума были получены значения из интервала [0,592 – 0,5965]. Отклонения от значения мета-характеристики без шума составили менее 1%, что характеризует мета-характеристику как устойчивую к шуму.

8.2. Шум-устойчивость модуля Noise Influence

Для тестирования мета-характеристики Noise Influence был использован временной ряд длиной в 3221 значение и с среднеквадратичным отклонением, равным 0.0270. Результатом вычисления на данном временном ряду является значение мета-характеристики, равное 1401,4.

Тестирование представляло собой 100 итераций с добавлением шума, представляющего собой случайные величины, распределенные по равномерному закону в интервале (0, 1) (рис 9).

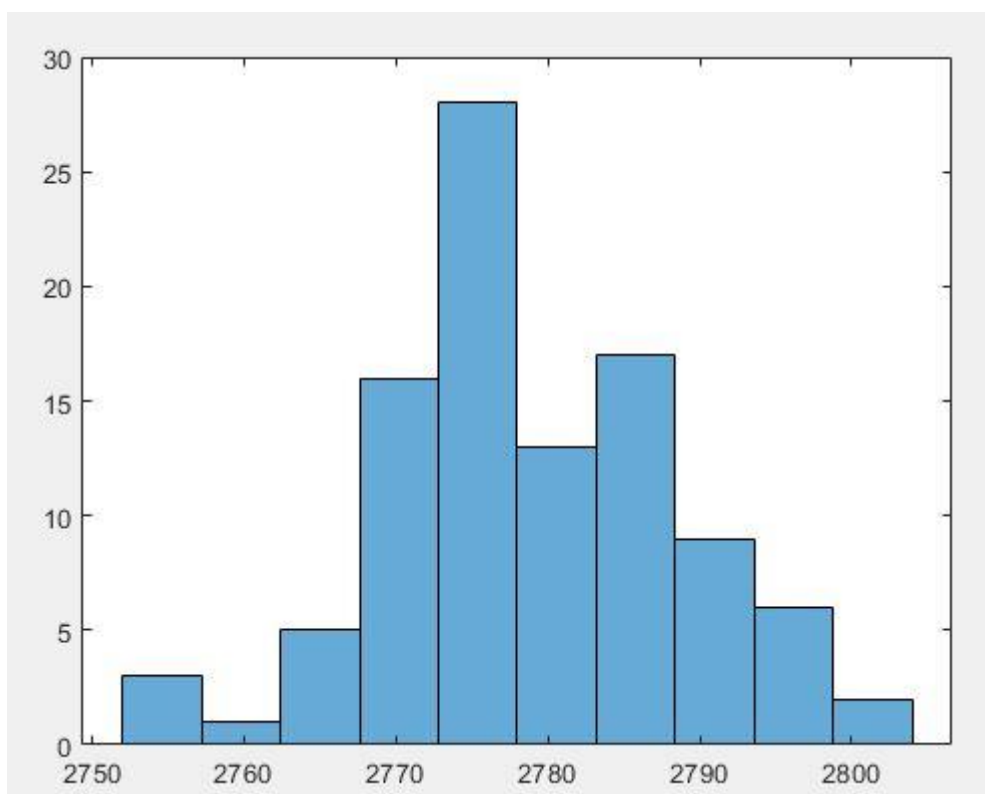


Рис 9. Гистограмма разброса значений после добавления шума

После добавления шума, значения увеличились более чем в 2 раза, что является естественным результатом, так как модуль предназначен для определения зашумленности ряда.

8.3. Шум-устойчивость модуля Chaos

Для тестирования мета-характеристики Chaos была использована система Лоренца [18] длиной в 5721 значение и с среднеквадратичным отклонением, равным 7.8616. Результатом вычисления на данном временном ряду является значение мета-характеристики, равное 0.6111.

Тестирование представляло собой 100 итераций с добавлением шума, представляющего собой случайные величины, распределенные по равномерному закону в интервале (0, 1) (рис 10).

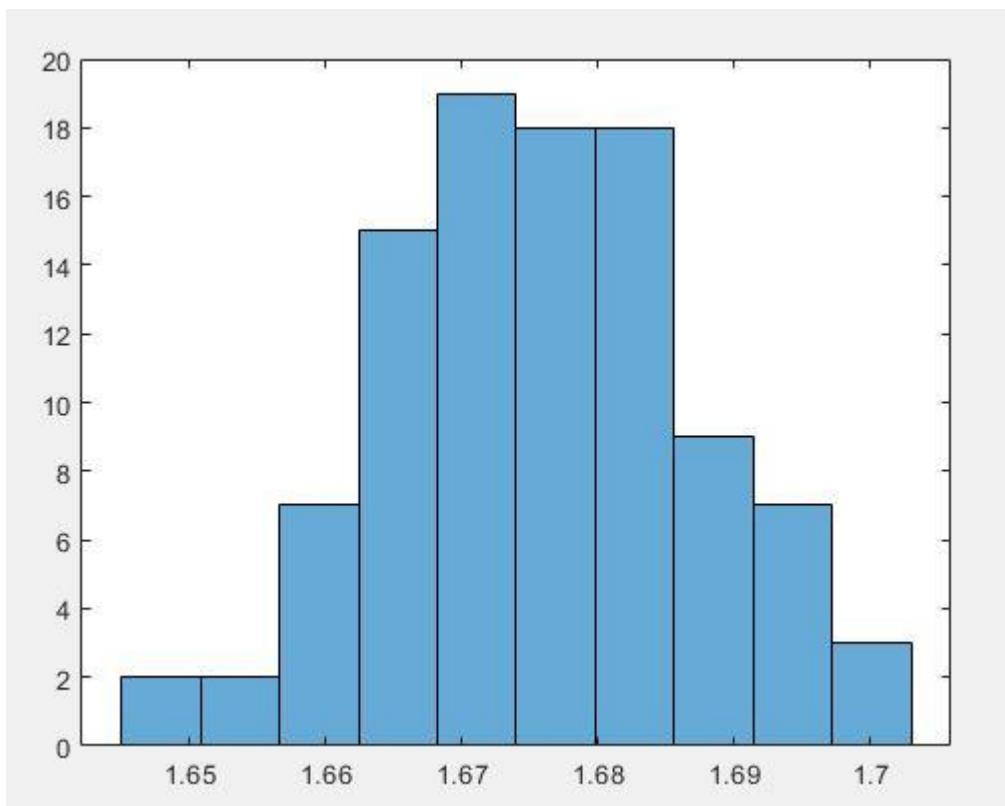


Рис 10. Гистограмма разброса значений после добавления шума

После добавления шума, значения увеличились, что является естественным результатом, и все еще характеризуют систему как слабо чувствительную к начальным условиям, из чего можно сделать вывод, что модуль устойчивый к шуму.

9. Финансовый менеджмент, ресурс эффективность и ресурсосбережение

Технико-экономическое обоснование научно-исследовательских работ проводится с целью определения и анализа трудовых и денежных затрат, направленных на их реализацию, а также уровня их научно-технической результативности.

Цель диссертации – разработка программных модулей для программной платформы, реализующей индуктивный подход к прогнозированию временных рядов.

Процесс разработки включает в себя этапы анализа, проектирования и создания модулей для системы. Необходимо произвести расчеты по данным этапам в соответствии с методическими требованиями.

9.1. Потенциальные потребители результатов исследования

Предприятие, по проблеме которого будет выполнена работа: общество с ограниченной ответственностью «Эко-Томск», г. Томск.

Для анализа потребителей результатов исследования необходимо рассмотреть целевой рынок и провести его сегментирование. **Целевой рынок** – сегменты рынка, на котором будет продаваться в будущем разработка. В свою очередь, сегмент рынка – это особым образом выделенная часть рынка, группы потребителей, обладающих определенными общими признаками. **Сегментирование** – это разделение покупателей на однородные группы, для каждой из которых может потребоваться определенный товар (услуга).

		Вид прогнозирования	
		Прогнозирование временных рядов с привлечением эксперта	Автоматизированное прогнозирование временных рядов
Размер компании	Крупные		
	Средние		
	Мелкие		

Таблица 2. Карта сегментирования рынка услуг по прогнозированию временных рядов:

Schlumberger
 Baker Hughes
 Weatherford

Судя по карте сегментирования можно сделать вывод, что ниша автоматизированного прогнозирования не занята на рынке. Исходя из этого, основным сегментом, на которое планируется ориентироваться, является незанятая ниша автоматизированного прогнозирования временных рядов.

9.1.1. Анализ конкурентных технических решений с позиции ресурс эффективности и ресурсосбережения

Детальный анализ конкурирующих разработок, существующих на рынке, необходимо проводить систематически, поскольку рынки пребывают в постоянном движении. Такой анализ помогает вносить коррективы в научное исследование, чтобы успешнее противостоять своим соперникам. Важно реалистично оценить сильные и слабые стороны разработок конкурентов.

Анализ конкурентных технических решений с позиции ресурс эффективности и ресурсосбережения позволяет провести оценку сравнительной эффективности научной разработки и определить направления для ее будущего повышения.

Таблица 3. Оценочная карта для сравнения конкурентных технических решений

Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
		Б _ф	Б _{к1}	Б _{к2}	К _ф	К _{к1}	К _{к2}
1	2	3	4	5	6	7	8
Технические критерии оценки ресурсоэффективности							
1. Повышение производительности труда пользователя	0,1	5	2	2	0,5	0,2	0,2
2. Удобство в эксплуатации (соответствует требованиям потребителей)	0,1	4	4	4	0,4	0,4	0,4
3. Помехоустойчивость	0,05	4	4	4	0,2	0,2	0,2
4. Энергоэкономичность	0,05	4	3	4	0,2	0,15	0,2
5. Надежность	0,05	4	4	3	0,2	0,2	0,15
6. Уровень шума	0,05	5	5	5	0,25	0,25	0,25
7. Безопасность	0,05	5	5	5	0,25	0,25	0,25
8. Потребность в ресурсах памяти	0,05	4	4	4	0,2	0,2	0,2
9. Функциональная мощность	0,05	5	3	3	0,25	0,15	0,15

(предоставляемые возможности)							
10. Простота эксплуатации	0,05	4	4	4	0,2	0,2	0,2
11. Качество интеллектуального интерфейса	0,02	4	4	4	0,08	0,08	0,08
12. Возможность подключения в сеть ЭВМ	0,05	5	3	4	0,25	0,15	0,2
Экономические критерии оценки эффективности							
1. Конкурентоспособность продукта	0,1	5	4	4	0,5	0,4	0,4
2. Уровень проникновения на рынок	0,03	4	4	4	0,12	0,12	0,12
3. Цена	0,02	4	4	3	0,08	0,08	0,06
4. Предполагаемый срок эксплуатации	0,05	5	3	4	0,25	0,15	0,2
5. Послепродажное обслуживание	0,03	4	4	4	0,12	0,12	0,12
6. Финансирование научной разработки	0,05	4	4	4	0,2	0,2	0,2
7. Срок выхода на рынок	0,03	4	4	4	0,12	0,12	0,12

8. Наличие сертификации разработки	0,02	4	4	4	0,08	0,08	0,08
Итого	1						

Вывод: анализ конкурентоспособности, представленный в таблице, показал, что технические и экономические показатели оценки ресурс эффективности нашей разработки опережают показатели конкурентов.

9.1.2. FAST-анализ

FAST-анализ выступает как синоним функционально-стоимостного анализа. Суть этого метода базируется на том, что затраты, связанные с созданием и использованием любого объекта, выполняющего заданные функции, состоят из необходимых для его изготовления и эксплуатации и дополнительных, функционально неоправданных, излишних затрат, которые возникают из-за введения ненужных функций, не имеющих прямого отношения к назначению объекта, или связаны с несовершенством конструкции, технологических процессов, применяемых материалов, методов организации труда и т.д.

Стадия 1. Объектом исследования является программная платформа.

Предметом исследования мета-характеристики, присутствующие и отсутствующие в данной программной платформе.

Стадия 2. Описание главной, основных и вспомогательных функций, выполняемых объектом. В рамках данной стадии FAST-анализа объект анализируется с позиции функционального устройства.

Таблица 4. Классификация функций, выполняемых объектом исследования

Наименование метода	Выполняемая функция	Ранг функции		
		Главная	Основная	Вспомогательная
Набор метаяхарактеристик	Подсчет метаяданных	X		
Ляпунов	Подсчет экспоненты Ляпунова		X	
Херст	Подсчет коэффициента Херста		X	
Разладка	Подсчет разладки		X	

Стадия 3. Определение значимости выполняемых функций объектом.

Для оценки значимости функций будем использовать метод расстановки приоритетов, предложенный Блумбергом В.А. и Глуценко В.Ф. В основу данного метода положено расчетно-экспертное определение значимости каждой функции.

Таблица 5. Матрица смежности

	Ляпунов	Херст	Разладка
Ляпунов	=	=	=
Херст	=	=	=
Разладка	=	=	=

Таблица 6. Матрица количественных соотношений функций

	Ляпунов	Херст	Разладка	Итого
Ляпунов	1	1	1	3
Херст	1	1	1	3
Разладка	1	1	1	3
				9

9.1.3. Диаграмма Исикава

Диаграмма причины-следствия Исикавы (Cause-and-Effect-Diagram) – это графический метод анализа и формирования причинно-следственных связей, инструментальное средство для систематического определения причин проблемы и последующего графического представления.

Область применения диаграммы:

- Выявление причин возникновения проблемы;
- Анализ и структурирование процессов на предприятии;
- Оценка причинно-следственных связей.

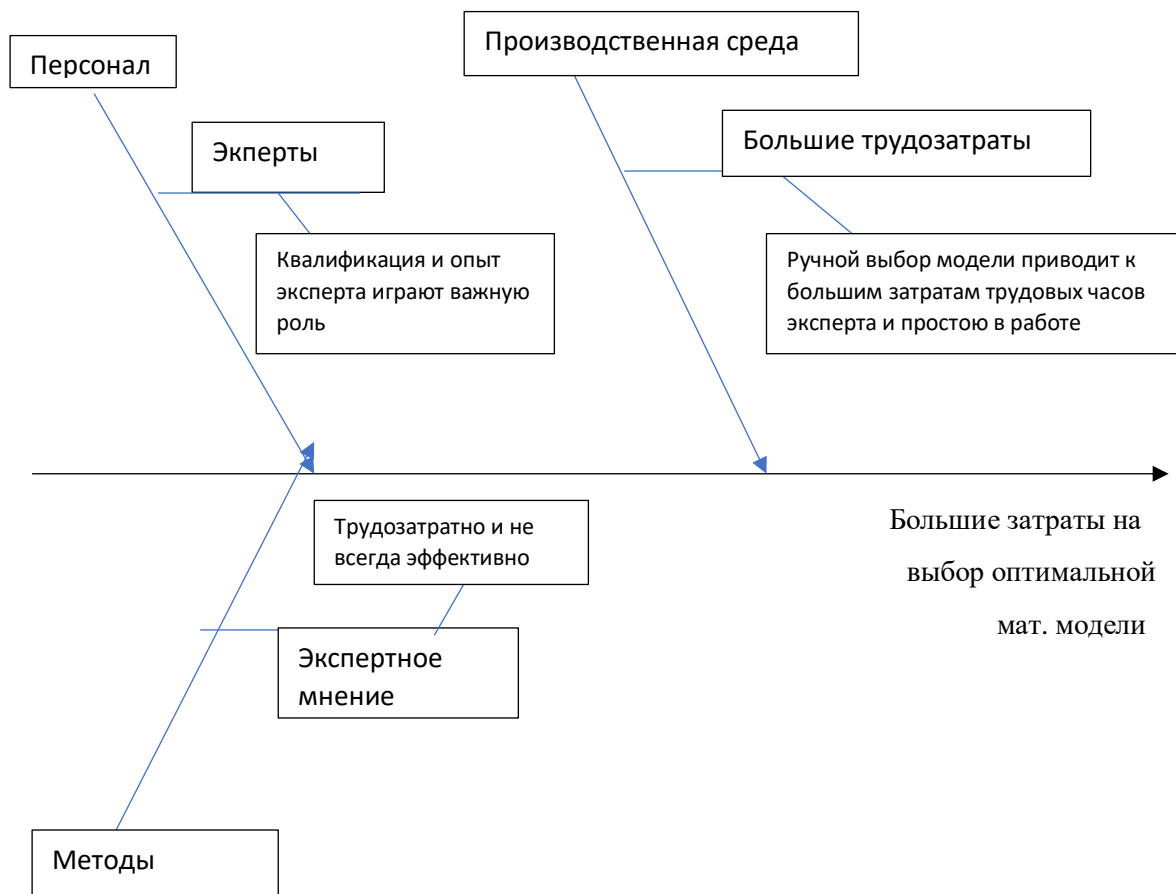


Диаграмма 1. Причинно-следственная диаграмма Исикавы

9.1.4. Технология Quad

Технология QuaD (QUality ADvisor) представляет собой гибкий инструмент измерения характеристик, описывающих качество новой разработки и ее перспективность на рынке и позволяющие принимать решение целесообразности вложения денежных средств в научно-исследовательский проект.

В основе технологии QuaD лежит нахождение средневзвешенной величины следующих групп показателей:

1) Показатели оценки коммерческого потенциала разработки:

- влияние нового продукта на результаты деятельности компании;
- перспективность рынка;
- пригодность для продажи;
- перспективы конструирования и производства;

- финансовая эффективность.
- правовая защищенность и др.

2) Показатели оценки качества разработки:

- динамический диапазон;
- вес;
- ремонтпригодность;
- энергоэффективность;
- долговечность;
- эргономичность;
- унифицированность;
- уровень материалоемкости разработки и др

Таблица 7. Оценочная карта для сравнения конкурентных технических решений

Критерии оценки	Вес критерия	Баллы	Максимальный балл	Относительное значение (3/4)	Средневзвешенное значение (5x2)
1	2	3	4	5	
Показатели оценки качества разработки					
1. Энергоэффективность	0,05	100	100	1	0,05
2. Помехоустойчивость	0,05	80	100	0,8	0,04
3. Надежность	0,05	90	100	0,9	0,045
4. Унифицированность	0,1	100	100	1	0,1
5. Уровень материалоемкости и разработки	0,05	80	100	0,8	0,04

6. Уровень шума	0,05	100	100	1	0,05
7. Безопасность	0,05	90	100	0,9	0,045
8. Потребность в ресурсах памяти	0,05	80	100	0,8	0,04
9. Функциональная мощность (предоставляемые возможности)	0,05	100	100	1	0,005
10. Простота эксплуатации	0,05	80	100	0,8	0,04
11. Качество интеллектуального интерфейса	0,02	80	100	0,8	0,016
12. Ремонтопригодность	0,05	90	100	0,9	0,045
Показатели оценки коммерческого потенциала разработки					
13. Конкурентоспособность продукта	0,1	100	100	1	0,1
14. Уровень проникновения на рынок	0,03	90	100	0,9	0,027
15. Перспективность рынка	0,05	90	100	0,9	0,045
16. Цена	0,02	90	100	0,9	0,018
17. Послепродажное обслуживание	0,03	90	100	0,9	0,027
18. Финансовая эффективность научной разработки	0,1	100	100	1	0,1
19. Срок выхода на рынок	0,03	80	100	0,8	0,024

20. Наличие сертификации разработки	0,02	90	100	0,9	0,018
ИТОГО	1				0,875

9.1.5. SWOT анализ

SWOT – Strengths (сильные стороны), Weaknesses (слабые стороны), Opportunities (возможности) и Threats (угрозы) – представляет собой комплексный анализ научно-исследовательского проекта. SWOT- анализ применяют для исследования внешней и внутренней среды проекта.

Результаты проведенного анализа отображены в таблице 8

Таблица 8. SWOT-анализ

	<p>Сильные стороны проекта:</p> <p>С1. Модульная структура. С2. Распределенность. Возможность распределять нагрузку между несколькими ЭВМ. С3. Регулируемая стоимость С4. Масштабируемость. Возможность внесения модификаций для расширения функционала. С5. Защищенность данных. С6. Высокое качество разрабатываемого продукта. С7. Поддержка и сопровождение продукта.</p>	<p>Слабые стороны проекта:</p> <p>Сл1. Малая популярность приложения. Сл2. Для каждого потребителя требуется индивидуальный подход. Сл3. Экономический спад Сл4. Смена спроса Сл5. Продукты-заменители</p>
<p>Возможности:</p> <p>В1. Расширение функционала приложения. В2. Увеличение рекламы продукта. В3. Появление дополнительного спроса на новый продукт. В4. Повышение стоимости конкурентных разработок.</p>	<p>С4В1 За счет масштабируемости есть возможность увеличить функционал программной платформы, что привлечет клиентов. С1С2С3С5С7В3 Модульная структура, распределенность, регулируемая стоимость, надежность вводимых данных, поддержка и</p>	<p>Сл1В1 Чтобы увеличить популярность приложения следует расширить функционал программной платформы и начать рекламную акцию. Сл3Сл5В4 Повышение стоимости аналогичных программных платформ конкурентов возможно по причине экономического спада, падения спроса или</p>

	сопровождение продукта привлекает клиентов. С6В2При транслировании рекламных роликов возможно увеличить количество пользователей. С3С6В4Если конкурентные компании повысят цену на продукт, клиентов привлечет низкая цена на наш продукт	возникновения более сильного конкурента
<p>Угрозы:</p> <p>У1. Отсутствие спроса на разрабатываемый продукт.</p> <p>У2. Появление более сильного конкурента.</p> <p>У3. Уменьшение целевой аудитории.</p> <p>У4. Изменение законодательства РФ.</p> <p>У5. Рост требований потребителей.</p>	<p>У1С1С3С5Избежать угрозы отсутствия или падения спроса, разработав программную платформу, позволяющую обрабатывать временные ряды в полностью автоматическом режиме прогнозировать данные с высокой точностью.</p> <p>Помимо этого стимуляцией спроса является снижение стоимости продукта.</p> <p>У2С3С6При появлении серьезного конкурента возможно урегулирование цены, расширение целевой аудитории.</p> <p>У3С3С7При уменьшении целевой аудитории следует снизить цену на продукт, выпустить обновление продукта.</p>	<p>У1Сл3При падении или исчезновении спроса на продукт возможно приостановление данного проекта.</p> <p>У2Сл2При появлении конкурента следует расширить опции программной платформы.</p> <p>У4Сл1При изменении законодательных актов РФ клиенты откажутся от пользования данной продукцией (например, если они будут вынуждены отказаться от ОС Windows).</p> <p>У5Сл2Сл3 Несвоевременная поддержка программной платформы может привести к снижению конкурентной позиций, а также экономическому спаду компании.</p>

В результате проведения анализа было решено в дальнейшем увеличить функционал программной платформы, чтобы привлечь новых клиентов (С4В1). Также было решено расширить опции программной платформы и постоянно выпускать обновления, чтобы избежать приостановления проекта (У2Сл2, У5Сл2Сл3).

9.2. Планирование управления научно-техническим проектом

9.2.1. Структура работ в рамках научного исследования

При организации процесса реализации конкретного проекта необходимо оптимально планировать занятость каждого из его участников и сроки проведения отдельных работ.

На данном этапе составляется полный перечень проводимых работ, и определяются их исполнители. Перечень работ указан в таблице 9.

Для выполнения перечисленных в таблице 9 работ требуются специалисты:

- студент – в его роли действуют исполнитель ВКР;
- научный руководитель.

Таблица 9. Перечень работ

Основные этапы	№ раб	Содержание работ	Должность исполнителя
Разработка технического задания	1	Согласование, утверждение и принятие технического задания	Научный руководитель
Исследовательская работа	2	Поиск новых метакarakterистик	Студент
	3	Установление подробного плана и сроков разработки модулей	Научный руководитель
	4	Утверждение технического проекта	Научный руководитель

Разработка алгоритмов, программной документации, испытание алгоритмов	5	Программирование, отладка модулей на локальной машине	Студент
	6	Разработка, выбор пакета тестов	Студент
	7	Поиск ошибок и уязвимостей	Студент
	8	Доработка проекта, устранение выявленных ошибок и уязвимостей	Студент
	9	Проверка на шум-устойчивость	Студент
Обобщение и оценка результатов	10	Перенос на Python	Студент
	11	Утверждение модулей	Научный руководитель
Оформление отчета по НИР (комплекта документации по ОКР)	12	Составление эксплуатационно-технической документации	Научный руководитель

9.2.2 Продолжительность этапов работ

Используем опытно-статистический метод, который реализуется двумя способами:

- аналоговый;
- вероятностный.

Для определения ожидаемого значения продолжительности работ тоже применяется вероятностный метод – метод двух оценок t_{\min} и t_{\max} .

$$t_{\text{ож}} = \frac{3*t_{\min} + 2*t_{\max}}{5} \quad (1)$$

где

t_{\min} – минимальная трудоемкость работ, чел/дн.;

t_{\max} – максимальная трудоемкость работ, чел/дн.

Рассчитаем длительность этапов в рабочих днях, а затем переведем в календарные дни. Формула расчета 2.

$$T_{РД} = \frac{t_{ож}}{K_{ВН}} * k_{Д} \quad (2)$$

где

$t_{ож}$ – трудоемкость работы, чел/дн.;

$K_{ВН}$ – коэффициент выполнения работ ($K_{ВН} = 1$);

$K_{Д}$ - коэффициент, учитывающий дополнительное время на компенсации и согласование работ ($K_{Д} = 1.2$).

Расчет продолжительности этапа в календарных днях ведется по формуле:

$$T_{КД} = T_{РД} * T_{К} \quad (3)$$

где

$T_{РД}$ – продолжительность выполнения этапа в рабочих днях;

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

$T_{К}$ – коэффициент календарности.

Коэффициент календарности рассчитывается по формуле:

$$T_{К} = \frac{T_{кал}}{T_{кал} - T_{вд} - T_{пд}} = \frac{365}{365 - 52 - 10} = 1,205 \quad (4)$$

где

$T_{КАЛ}$ – календарные дни ($T_{КАЛ} = 365$);

$T_{ВД}$ – выходные дни ($T_{ВД} = 52$);

$T_{ПД}$ – праздничные дни ($T_{ПД} = 10$).

В таблице 10 приведены длительности этапов работ и число исполнителей, занятых на каждом этапе, в таблице 11 линейный график работ по выполнению ВКР.

Таблица 10. Трудозатраты на выполнение проекта

Этап	Исполнители	Продолжительность работ, дни			Длительность работ, чел/дн.			
		t_{min}	t_{max}	$t_{ож}$	$T_{рд}$		$T_{кд}$	
					НР	И	НР	И
1	2	3	4	5	6	7	8	9
Согласование, утверждение и принятие технического задания	НР	1,00	3,00	1,80	2,16	-	2,60	-
Поиск новых мета-характеристик	С	10,00	20,00	14,00	-	16,80	-	20,24
Установление подробного плана и сроков разработки модулей	НР	3,00	5,00	3,80	4,56	-	5,50	-
Утверждение технического проекта	НР	4,00	6,00	4,80	5,76	-	6,94	-
Программирование, отладка модулей на локальной машине	С	15,00	20,00	17,00	-	20,40	-	24,59
Разработка, выбор пакета тестов	С	10,00	14,00	11,60	-	13,92	-	16,77
Поиск ошибок и уязвимостей	С	7,00	10,00	8,20	-	9,84	-	11,86
Доработка проекта, устранение выявленных ошибок и уязвимостей	С	10,00	20,00	10,00	-	12,00	-	14,46
Проверка на шум-устойчивость	С	6,00	9,00	7,20	-	8,64	-	10,41
Перенос на Python	С	5,00	6,00	5,40	-	6,48	-	7,80
Утверждение модулей	НР	3,00	5,00	3,80	4,56	-	5,50	-
Составление эксплуатационно-технической документации	НР	2,00	3,00	2,40	2,88	-	3,47	-
Итого:				90,00	19,92	88,08	24,00	106,14

Таблица 11. Линейный график работ

Этап	НР	С	Март			Апрель			Май			Июнь	
			10	20	30	40	50	60	70	80	90	100	110
1	2,60	-	■										
2	-	20,24	■	■									
3	5,50	-			■								
4	6,94	-			■	■							
5	-	24,59			■	■	■						
6	-	16,77					■	■					
7	-	11,86							■	■			
8	-	14,46								■	■		
9	-	10,41										■	■
10	-	7,80											■
11	5,50	-										■	■
12	3,47	-											■

■ - Научный руководитель
 ■ - Студент

9.2.3 Расчет накопления готовности проекта

В данном пункте будет описан расчет текущих состояний работы над проектом. Величина накопления готовности работы показывает, на сколько процентов по окончании текущего этапа выполнен общий объем работ по проекту в целом.

Введем обозначения:

$TP_{\text{общ}}$ – общая трудоемкость проекта;

TP_i (TP_k) – трудоемкость i -го (k -го) этапа проекта;

TP_i^H – накопленная трудоемкость i -го этапа проекта;

TP_{ij} (TP_{kj}) – трудоемкость работ, выполняемых j -м участником на i -м этапе, здесь $j = (1, m)$ индекс исполнителя, в нашем примере $m = 2$.

Степень готовности определяется формулой 5

$$CG_i = \frac{TP_i^H}{TP_{\text{общ}}} = \frac{\sum_{k=1}^i TP_k}{TP_{\text{общ}}} = \frac{\sum_{k=1}^i \sum_{j=1}^m TP_{km}}{\sum_{k=1}^i \sum_{j=1}^m TP_{km}} \quad (5)$$

Таблица 12. Нарастание технической готовности работы и удельный вес каждого этапа

Этап	TP_i , %	CG_i , %
Согласование, утверждение и принятие технического задания	2	2
Поиск новых мета-характеристик	15,55	17,55
Установление подробного плана и сроков разработки модулей	4,22	21,77
Утверждение технического проекта	5,33	27,1
Программирование, отладка модулей на локальной машине	18,90	46
Разработка, выбор пакета тестов	12,89	58,89
Поиск ошибок и уязвимостей	9,11	68
Доработка проекта, устранение выявленных ошибок и уязвимостей	11,11	79,11
Проверка на шум-устойчивость	8	87,11
Перенос на Python	6	93,11
Утверждение модулей	4,22	97,33
Составление эксплуатационно-технической документации	2,67	100

9.2.4 Расчет сметы затрат на выполнение проекта

Расчет сметной стоимости выполнения проекта будет производиться по следующим статьям затрат:

- материалы и покупные изделия;
- заработная плата;
- социальный налог;
- расходы на электроэнергию (без освещения);
- амортизационные начисления;
- оплата услуг связи;
- прочие (накладные расходы) расходы.

9.2.4.1 Расчет затрат на материалы

К данной статье расходов относится стоимость материалов, покупных изделий, полуфабрикатов и других материальных ценностей, расходуемых непосредственно в процессе выполнения работ над объектом проектирования. Цена материальных ресурсов определяется по соответствующим ценникам или договорам поставки. Сюда же включаются расходы на совершение сделки купли-продажи. Приблизительно они оцениваются в процентах к отпускной цене закупаемых материалов, как правило, это 5 ÷ 20 %. Исполнитель работы самостоятельно выбирает их величину в указанных границах.

Таблица 13 – Расчет затрат на материалы

Наименование материалов	Цена за ед., руб.	Кол-во	Сумма, руб.
Бумага для принтера формата А4	210	2 шт.	420
Заправка картриджа	500	2 шт.	1000
Итого:			1420

Допустим, что ТЗР (транспортно-заготовительные расходы) составляют 5 % от отпускной цены материалов, тогда расходы на материалы с учетом ТЗР равны $S_{\text{МАТ}} = 1491$.

9.2.4.2 Расчет заработной платы

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

$$ЗП_{дн-т} = MO/N \quad (6)$$

N – количество рабочих дней в месяц, при шестидневной рабочей неделе – $N = 24.91$, а при пятидневной рабочей неделе – 20.58 .

Учитывая, что в году 299 рабочих дней и, следовательно, в месяце в среднем 24,91 (299/12) рабочих дня при шестидневной рабочей неделе.

Расчеты затрат на полную заработную плату приведены в таблице 14. Затраты времени по каждому исполнителю в рабочих днях с округлением до целого взяты из таблицы 10. Для учета в ее составе премий, дополнительной зарплаты и районной надбавки используется следующий ряд коэффициентов: $K_{ПР} = 1,1$; $K_{доп.ЗП} = 1,188$; $K_p = 1,3$. Таким образом, для перехода от тарифной (базовой) суммы заработка исполнителя, связанной с участием в проекте, к соответствующему полному заработку (зарплатной части сметы) необходимо первую умножить на интегральный коэффициент $K_{и} = 1,1 * 1,188 * 1,3 = 1,699$. Вышеуказанное значение $K_{доп.ЗП}$ применяется при шестидневной рабочей неделе, при пятидневной оно равно 1,113, соответственно в этом случае $K_{и} = 1,62$.

Таблица 14. Затраты на основную заработную плату

Исполнитель	Оклад, руб./мес.	Среднедневная ставка, руб./день	Затраты времени, дни	Коэффициент	Фонд з/платы, руб.
НР	30 000	1205	20	1,699	40946
С	15 000	730	89	1,62	105251
Итого:					146197

Затраты на заработную плату составили $COCH = 146197$ руб.

9.2.4.3 Расчет затрат на социальный налог

Затраты на единый социальный налог (ЕСН), включающий в себя отчисления в пенсионный фонд, на социальное и медицинское страхование, составляют 30,2% от полной заработной платы по проекту, т.е. $C_{соц.} = C_{зп} * 0,302$. Итак, в нашем случае $C_{соц.} = 146197 * 0,302 = 44151,5$ руб.

9.2.4.4 Расчет затрат на электроэнергию

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

$$C_{эл.об} = P_{об} * t_{об} * Ц_{э} \quad (7)$$

где

$P_{об}$ – мощность, потребляемая оборудованием, кВт;

$Ц_{э}$ – тариф на 1 кВт·час;

$t_{об}$ – время работы оборудования, час.

Для ТПУ $Ц_{э} = 5,257$ руб./кВт·час (с НДС).

Время работы оборудования вычисляется на основе итоговых данных таблицы 12 для студента (ТРД) из расчета, что продолжительность рабочего дня равна 8 часов.

$$t_{об} = T_{РД} * K_t \quad (8)$$

где $K_t \leq 1$ – коэффициент использования оборудования по времени, равный отношению времени его работы в процессе выполнения проекта к $T_{РД}$, определяется исполнителем самостоятельно (0,9). В ряде случаев возможно определение $t_{об}$ путем прямого учета, особенно при ограниченном использовании соответствующего оборудования.

Мощность, потребляемая оборудованием, определяется по формуле:

$$P_{об} = P_{ном} * K_c \quad (9)$$

где

$P_{ном}$ – номинальная мощность оборудования, кВт;

$K_c \leq 1$ – коэффициент загрузки, зависящий от средней степени использования номинальной мощности. Для технологического оборудования малой мощности $K_c = 1$.

Пример расчета затраты на электроэнергию для технологических целей приведен в таблице 15.

Таблица 15. Затраты на электроэнергию технологическую

Наименование оборудования	Время работы оборудования $t_{ОБ}$, час	Потребляемая мощность РОБ, кВт	Затраты ЭОБ, руб.
Персональный компьютер	626,4	0,08	289,74
Струйный принтер	30	0,1	15,77
Итого:			305,51

9.2.4.5 Расчет амортизационных расходов

$$C_{AM} = \frac{N_A * C_{ОБ} * t_{рф} * n}{F_d} \quad (10)$$

где N_A – годовая норма амортизации единицы оборудования.

$C_{ОБ}$ – балансовая стоимость единицы оборудования с учетом ТЗР.

Стоимость ПК студента – 20000 руб.;

F_d – действительный годовой фонд времени работы соответствующего оборудования, берется из специальных справочников или фактического режима его использования в текущем календарном году. При этом второй вариант позволяет получить более объективную оценку C_{AM} . Например,

для ПК в 2017 г. (299 рабочих дней при шестидневной рабочей неделе) можно принять $F_D = 299 * 8 = 2392$ часа;

$t_{рф}$ – фактическое время работы оборудования в ходе выполнения проекта, учитывается исполнителем проекта

$$t_{рф} = 64,96 * 8 = 626,4$$

N_A определяется по формуле:

$$N_A = \frac{1}{C_A} \quad (11)$$

где C_A – срок амортизации, который можно получить из постановления правительства РФ «О классификации основных средств, включенных в амортизационные группы» Для электронно-вычислительной техники C_A свыше 2 лет до 3 лет включительно. В данной работе примем $C_A=2,5$ года. Тогда

$$N_A = \frac{1}{2,5} = 0,4$$

n – число задействованных однотипных единиц оборудования.

Таким образом,

$$C_{AM} = \frac{0,4 * 20000 * 626,4 * 1}{2392} = 2095 \text{ руб}$$

9.2.4.6 Расчет расходов на услуги связи

Расходы на услуги связи определены наличием подключения к сети Интернет на компьютерах, использованных в данной работе.

Ежемесячная оплата, согласно тарифу TRU net, составляет 350 рублей. В соответствии с таблицей 11, трудоемкость выполняемой задачи составляет четыре календарных месяца. Таким образом, сумма расходов на услуги связи составляет $4*350 = 1400$ руб. Общая сумма расходов $C_{св} = 1400$ руб.

9.2.4.7 Расчет прочих расходов

В статье «Прочие расходы» отражены расходы на выполнение проекта, которые не учтены в предыдущих статьях, их следует принять равными 10% от суммы всех предыдущих расходов, т.е.

$$C_{\text{проч}} = (1491 + 146197 + 44151,5 + 305,51 + 2095 + 1400) * 0,1 = 19564$$

9.2.4.8 Расчет общей себестоимости разработки

Смета на разработку проекта приведена в таблице 16.

Таблица 16. Затраты на разработку проекта

Статья затрат	Условное обозначение	Сумма, руб.
Материалы и покупные изделия	С _{мат}	1491
Основная заработная плата	С _{ЗП}	146197
Отчисления в социальные фонды	С _{соц}	44151,5
Расходы на электроэнергию	С _{эл.об.}	305,51
Амортизационные отчисления	С _{ам}	2095
Расходы на услуги связи	С _{св}	1400
Прочие расходы	С _{проч}	19564
Итого:		215204

Таким образом, затраты на разработку составили $C = 215204$ руб.

9.2.4.9 Прибыль

Данная работа является социально направленной и не представляет прямой экономической выгоды.

Прибыль составляет 43040,8 рублей (20 %) от расходов на разработку.

9.2.4.10 НДС

НДС составляет 18% от суммы затрат на разработку и прибыли. В нашем случае это $(215204 + 43040,8) * 0,18 = 46484,06$ руб.

9.2.4.11 Цена разработки НИР

Цена равна сумме полной себестоимости, прибыли и НДС, в нашем случае $C_{\text{НИР(КР)}} = 215204 + 43040,8 + 46484,06 = 304728,86$ руб.

9.3 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

Определение эффективности происходит на основе расчета интегрального показателя эффективности научного исследования. Его нахождение связано с определением двух средневзвешенных величин: финансовой эффективности и ресурсоэффективности.

Интегральный показатель финансовой эффективности научного исследования получают в ходе оценки бюджета затрат трех (или более) вариантов исполнения научного исследования (см. табл. 17). Для этого наибольший интегральный показатель реализации технической задачи принимается за базу расчета (как знаменатель), с которым соотносятся финансовые значения по всем вариантам исполнения.

Интегральный финансовый показатель разработки определяется как:

$$I_{\text{финр}}^{\text{исп.}i} = \frac{\Phi_{pi}}{\Phi_{\text{max}}} \quad (12)$$

где $I_{\text{финр}}^{\text{исп.}i}$ - интегральный финансовый показатель разработки;

Φ_{pi} – стоимость i -го варианта исполнения;

Φ_{max} – максимальная стоимость исполнения научно-исследовательского проекта (в т.ч. аналоги).

Полученная величина интегрального финансового показателя разработки отражает соответствующее численное увеличение бюджета затрат разработки в размах (значение больше единицы), либо соответствующее численное удешевление стоимости разработки в размах (значение меньше единицы, но больше нуля).

Интегральный показатель ресурсоэффективности вариантов исполнения объекта исследования можно определить следующим образом:

$$I_{pi} = \sum a_i * b_i \quad (13)$$

где I_{pi} – интегральный показатель ресурсоэффективности для i -го варианта исполнения разработки;

a_i – весовой коэффициент i -го варианта исполнения разработки;

b_i^a , b_i^p – бальная оценка i -го варианта исполнения разработки, устанавливается экспертным путем по выбранной шкале оценивания;

n – число параметров сравнения.

Таблица 17. Сравнительная оценка характеристик вариантов исполнения проекта

Критерий	Весовой коэффициент параметра	Исп. 1	Исп. 2	Исп. 3
1. Способствует росту производительности труда пользователя	0,5	5	3	4
2. Удобство в эксплуатации (соответствует требованиям потребителей)	0,05	4	4	4
3. Помехоустойчивость	0,15	4	4	4
4. Энергосбережение	0,05	4	4	3
5. Надежность	0,2	4	4	4
6. Материалоемкость	0,05	5	4	4
ИТОГО	1			

Интегральный показатель эффективности вариантов исполнения разработки ($I_{исп.i}$) определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формуле:

$$I_{исп.i} = \frac{I_{р-исп.i}}{I_{финр}^{исп.i}} \quad (14)$$

Сравнение интегрального показателя эффективности вариантов исполнения разработки позволит определить сравнительную эффективность проекта (см.табл.18) и выбрать наиболее целесообразный вариант из предложенных. Сравнительная эффективность проекта ($\mathcal{E}_{ср}$):

$$\mathcal{E}_{ср} = \frac{I_{исп1}}{I_{исп2}} \quad (15)$$

Таблица 18. Сравнительная эффективность разработки

№ п/п	Показатели	Исп. 1	Исп. 2	Исп. 3
1	Интегральный финансовый показатель разработки	0,89	1,00	0,95
2	Интегральный показатель ресурс эффективности разработки	4,55	3,5	3,95
3	Интегральный показатель эффективности	5,11	3,5	4,16
4	Сравнительная эффективность вариантов исполнения	1,46 / 1,23	0,69 / 0,84	0,81 / 1,19

Вывод: сравнение значений интегральных показателей эффективности позволило выбрать наиболее эффективный вариант решения проекта. С

позиций технической и финансовой ресурсоэффективности мы можем сделать выводы о том, что научно – техническое решение, представленное первым исполнителем, является более предпочтительным.

10 Социальная ответственность

Во время разработки и эксплуатации проекта «Анализ индуктивного метода прогнозирования временных рядов и поиск мета-характеристик для улучшения его репрезентативности» выполнялись работы, связанные построением алгоритмов, моделированием процессов, анализом информации, программированием. Программный комплекс будет использоваться для анализа и прогнозирования временных рядов.

Предприятие, по проблеме которого будет выполнена работа: ООО «Эко-Томск», г. Томск. В связи с тем, что работа связана непосредственно с компьютером, то такая работа может отрицательно воздействовать на здоровье человека. Во-первых, большую угрозу несёт монитор компьютера, так как он является источником электромагнитного поля. Во-вторых, неподвижная напряжённая поза в течение продолжительного временного периода способствует к быстрому переутомлению и появлению болевых ощущений в области позвоночника, плечевых суставов, шеи. Однако, наиболее сильной нагрузке подвергаются глаза человека. Ещё один немаловажный фактор – это работа с клавиатурой. Данный вид работы вызывает боль в локтевых суставах, запястьях, кистях и пальцах рук.

10.1 Производственная безопасность

В таблице 19 представлены основные виды работ, которые могут привести к воздействию опасных и вредных факторов.

Таблица 19. Основные элементы производственного процесса, формирующие опасные и вредные факторы рабочего места разработчика

Наименование видов работ	Факторы (по ГОСТ 12.0.003 - 74)		Нормативные документы
	Вредные	Опасные	
Работа за персональным	1) повышенный уровень электромагнитн	1) Повышенный уровень статического	Работа за персональным

компьютером в офисном помещении	ых излучений; 2) повышенная напряжённость электрического поля;	электричества; пожароопасность.	компьютером в офисном помещении
---------------------------------------	--	------------------------------------	---------------------------------------

10.1.1 Повышенный уровень электромагнитных излучений; повышенная напряжённость электрического поля

Источником электромагнитного поля и электромагнитных излучений на рабочем месте является компьютер, в частности, экран монитора компьютера. Электромагнитное поле, создаваемое ПК, имеет сложный спектральный состав в диапазоне частот от 0 Гц до 1000 МГц, и в том числе мощность экспозиционной дозы мягкого рентгеновского излучения в любой точке на расстоянии 0,05 м от экрана при любых положениях ПК не должна превышать 100 мкР/час.

Время работы на ПК по санитарным нормам не должно превышать четыре часа.

Допустимые значения параметров неионизирующих электромагнитных излучений от монитора компьютера представлены в таблице 20.

Таблица 20. Допустимые значения параметров неионизирующих электромагнитных излучений ПО ГОСТ 12.1.045–84

Наименование параметра	Допустимые значения
Напряженность электрической составляющей электромагнитного поля на расстоянии 50см от поверхности видеомонитора	10 В/м
Напряженность магнитной	0,3 А/м

составляющей электромагнитного поля на расстоянии 50см от поверхности видеомонитора	
Напряженность электростатического поля не должна превышать: - для взрослых пользователей - для детей дошкольных учреждений и учащихся средних специальных и высших учебных заведений	20 кВ/м 15 кВ/м

Предельно-допустимые нормы ЭМП представлены в таблице 21.

Таблица 21. Предельно допустимые нормы ЭМП ПО ГОСТ 12.1.045–84

Напряжённость электрического поля	
в диапазоне частот 5 Гц–2 кГц	в диапазоне частот 5 Гц–2 кГц
в диапазоне частот 2 кГц–400 кГц	в диапазоне частот 2 кГц–400 кГц
Плотность магнитного потока	
в диапазоне частот 5 Гц–2 кГц	в диапазоне частот 5 Гц–2 кГц
в диапазоне частот 2 кГц–400 кГц	в диапазоне частот 2 кГц–400 кГц

Ряд мероприятий, позволяющих уменьшить влияние вредных факторов на работника при работе за ПК: каждый час необходимо делать перерыв, для выполнения гимнастики для глаз, а также выполнять несколько упражнений на расслабление, которые могут уменьшить напряжение, накапливающееся в мышцах при длительной работе за компьютером.

Основные способы защиты от статического электричества следующие: заземление оборудования, увлажнение окружающего воздуха. Также целесообразно применение полов из антистатического материала.

10.1.2 Электрический ток (источник: ПК)

Токи статического электричества, наведённые в процессе работы компьютера на корпусах монитора, системного блока, клавиатуры, а также при работе за паяльной станцией, могут приводить к разрядам при прикосновении к этим элементам. Такие разряды опасности для человека не представляют, но могут привести к выходу из строя вышеописанного оборудования.

На рабочем месте пользователя размещены дисплей, клавиатура и системный блок. Использование паяльной станции со всеми необходимыми принадлежностями предполагается отдельно, когда на рабочем месте могут присутствовать все элементы ПК, но они находятся на расстоянии не менее вытянутой руки сидящего работника (по ГОСТ 12.1.038–82).

Перед началом работы следует убедиться в отсутствии свешивающихся со стола или висящих под столом проводов электропитания, в целостности вилки и провода электропитания, в отсутствии видимых повреждений аппаратуры и рабочей мебели, в отсутствии повреждений и наличии заземления приэкранного фильтра.

Методы защиты от воздействия статического электричества (по ГОСТ 12.1.038–82):

- влажная уборка, чтобы уменьшить количество пылинок в воздухе и на предметах офиса;
- использование увлажнителей воздуха;
- защитное заземление;
- применение средств индивидуальной защиты, таких как антистатические спреи и браслеты.

Допустимый ток частотой 50 Гц при длительности воздействия более 10 секунд составляет 2 мА, а при длительности 10 секунд и менее – 6 мА. Для переменного тока эта величина соответственно равна 10 и 15 мА.

Методы защиты от опасности поражения электрическим током (по ГОСТ 12.1.038–82):

- электрическая изоляция токоведущих частей (сопротивление изоляции должно быть не менее 0,5 МОм);
- ограждение токоведущих частей, которые работают под напряжением;
- использование малых напряжений, например, не более 50 В;
- электрическое разделение сетей на отдельные короткие участки;
- защитное заземление и зануление;
- применение средств индивидуальной защиты, таких как плакаты и знаки безопасности, изолирующие подставки, указатели напряжения.

10.2 Экологическая безопасность

В данном разделе рассматривается воздействие на окружающую среду деятельности по разработке проекта.

При разработке используется ПК, который потребляет сравнительно небольшое количество электроэнергии (мощностью 220 Вт).

Под охраной окружающей среды характеризуется различного рода мероприятиями, влияющие на следующие природные зоны: атмосфера, гидросфера, литосфера.

10.2.1 Анализ воздействия на окружающую среду

Охрана окружающей среды - деятельность органов государственной власти Российской Федерации, органов государственной власти субъектов Российской Федерации, органов местного самоуправления, общественных объединений и некоммерческих организаций, юридических и физических лиц, направленная на сохранение и восстановление природной среды, рациональное использование и воспроизводство природных ресурсов, предотвращение негативного воздействия хозяйственной и иной деятельности на окружающую среду и ликвидацию ее последствий. Наиболее активной формой защиты окружающей среды от вредного воздействия выбросов

промышленных предприятий является полный переход к безотходным, малоотходным и энергосберегающим технологиям и производствам.

В ходе выполнения ВКР и дальнейшем использовании алгоритмов отсутствуют выбросы каких-либо вредных веществ в атмосферу, следовательно, загрязнение воздуха не происходит. Не происходит также и сбросов в водоемы, поэтому не оказывается никакого влияния на гидросферу. В процессе разработки возникает необходимость утилизировать производственные отходы, в качестве которых в данном случае выступают бумажные отходы (макулатура) и неисправные детали персональных компьютеров, плат, контроллеров.

10.3 Безопасность в чрезвычайных ситуациях

В данном разделе рассматривается возможность возникновения чрезвычайных ситуаций при разработке и применении проекта.

Так как при работе используется ПК, находящийся в компьютерной аудитории или помещении, такая чрезвычайная ситуация как возникновение пожара имеет наибольшую вероятность возникновения.

10.3.1 Пожарная безопасность

Компьютерный класс по пожарной безопасности относится к категории В, в нём находятся горючие материалы и вещества в холодном состоянии (по ГОСТ 12.1.010-76). По степени огнестойкости данное помещение относится к 3-й степени огнестойкости (по ГОСТ 12.1.004-91). Возможные причины пожара: перегрузка в электросети, короткое замыкание, разрушение изоляции проводников.

Для локализации или ликвидации загорания на начальной стадии используются первичные средства пожаротушения (по ГОСТ 12.1.004-91):

- огнетушащие вещества (вода, песок, земля);
- огнетушащие материалы (грубошёрстные куски материи – кошмы, асбестовые полотна, металлические сетки с малыми ячейками ит. п.);

- немеханизированный ручной пожарный инструмент (багры, крюки, ломы, лопаты и т.п.);
- пожарный инвентарь (бочки и чаны с водой, пожарные ведра, ящики и песочницы с песком);
- пожарные краны на внутреннем водопроводе противопожарного водоснабжения в сборе с пожарным стволом и пожарным рукавом;
- огнетушители.

Первичные средства пожаротушения обычно применяют до прибытия пожарной команды.

Здание должно соответствовать требованиям пожарной безопасности, а именно: наличие охранно-пожарной сигнализации, плана эвакуации, порошковых огнетушителей с поверенным клеймом, табличек с указанием направления к запасному (эвакуационному) выходу.

Углекислотные огнетушители ОУ-3, ОУ-5 предназначены для тушения загораний веществ, горение которых не может происходить без доступа воздуха, загораний электроустановок, находящихся под напряжением не более 1000 В, жидких и газообразных веществ (класс В, С).

Огнетушители не предназначены для тушения загорания веществ, горение которых может происходить без доступа воздуха (алюминий, магний и их сплавы, натрий, калий), такими огнетушителями нельзя тушить дерево.

В общественных зданиях и сооружениях на каждом этаже должно размещаться не менее двух переносных огнетушителей. Огнетушители следует располагать на видных местах вблизи от выходов из помещений на высоте не более 1,35 м. Размещение первичных средств пожаротушения в коридорах, переходах не должно препятствовать безопасной эвакуации людей (по ГОСТ 12.1.004-91).

10.4 Правовые и организационные вопросы обеспечения безопасности

10.4.1 Организационные мероприятия обеспечения безопасности

Предъявляемые требования к расположению и компоновке рабочего места: «Высота рабочей поверхности стола для взрослых пользователей должна регулироваться в пределах (680÷800) мм, при отсутствии такой возможности высота рабочей поверхности стола должна составлять 725 мм» (по СанПиН 2.2.2/2.4.1340–03).

Рабочий стол должен иметь пространство для ног высотой не менее 600 мм, шириной – не менее 500 мм, глубиной на уровне колен – не менее 450 мм и на уровне вытянутых ног – не менее 650 мм (по СанПиН 2.2.2/2.4.1340–03).

Конструкция рабочего стула должна обеспечивать (по СанПиН 2.2.2/2.4.1340–03):

- ширину и глубину поверхности сиденья не менее 400 мм;
- поверхность сиденья с закругленным передним краем;
- регулировку высоты поверхности сиденья в пределах (400÷550) мм и углам наклона вперед до 15 град, и назад до 5 град.;
- высоту опорной поверхности спинки (300±20) мм, ширину – не менее 380 мм и радиус кривизны горизонтальной плоскости –400 мм;
- угол наклона спинки в вертикальной плоскости в пределах ±30 градусов;
- регулировку расстояния спинки от переднего края сиденья в пределах (260÷400) мм;
- стационарные или съемные подлокотники длиной не менее 250 мм и шириной –(50÷70) мм;
- регулировку подлокотников по высоте над сиденьем в пределах (230±30) мм и внутреннего расстояния между подлокотниками в пределах (350÷500) мм.

Рабочее место пользователя ПК следует оборудовать подставкой для ног, имеющей ширину не менее 300 мм, глубину не менее 400 мм, регулировку по высоте в пределах до 150 мм и по углу наклона опорной поверхности подставки до 20°. Поверхность подставки должна быть рифленой и иметь по переднему краю бортик высотой 10 мм (по СанПиН 2.2.2/2.4.1340–03).

10.4.2 Правовые мероприятия обеспечения безопасности

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.

Порядок исчисления нормы рабочего времени на определенные календарные периоды (месяц, квартал, год) в зависимости от установленной продолжительности рабочего времени в неделю определяется федеральным органом исполнительной власти (утв. Приказом Минздравсоцразвития РФ от 13.08.2009 N 588н.), осуществляющим функции по выработке государственной политики и нормативно-правовому регулированию в сфере труда.

Нормальная продолжительность рабочего времени не может превышать 40 часов в неделю.

Из статьи 92 ТК РФ:

Сокращенная продолжительность рабочего времени устанавливается:

- для работников в возрасте до шестнадцати лет - не более 24 часов в неделю;
- для работников в возрасте от шестнадцати до восемнадцати лет - не более 35 часов в неделю;
- для работников, являющихся инвалидами I или II группы, - не более 35 часов в неделю;

- для работников, занятых на работах с вредными и (или) опасными условиями труда, - не более 36 часов в неделю в порядке, установленном Правительством Российской Федерации с учетом мнения Российской трехсторонней комиссии по регулированию социально-трудовых отношений.

Продолжительность рабочего времени учащихся образовательных учреждений в возрасте до восемнадцати лет, работающих в течение учебного года в свободное от учебы время, не может превышать половины норм, установленных частью первой настоящей статьи для лиц соответствующего возраста.

Продолжительность ежедневной работы (смены) не может превышать:

- для работников в возрасте от пятнадцати до шестнадцати лет - 5 часов, в возрасте от шестнадцати до восемнадцати лет - 7 часов;
- для учащихся общеобразовательных учреждений, образовательных учреждений начального и среднего профессионального образования, совмещающих в течение учебного года учебу с работой, в возрасте от четырнадцати до шестнадцати лет - 2,5 часа, в возрасте от шестнадцати до восемнадцати лет - 4 часов;
- для инвалидов - в соответствии с медицинским заключением, выданным в порядке, установленном федеральными законами и иными нормативными правовыми актами Российской Федерации.

Для работников, занятых на работах с вредными и (или) опасными условиями труда, где установлена сокращенная продолжительность рабочего времени, максимально допустимая продолжительность ежедневной работы (смены) не может превышать:

- при 36-часовой рабочей неделе - 8 часов;
- при 30-часовой рабочей неделе и менее - 6 часов.

В течение рабочего дня (смены) работнику должен быть предоставлен перерыв для отдыха и питания продолжительностью не более двух часов и не менее 30 минут, который в рабочее время не включается.

Время предоставления перерыва и его конкретная продолжительность устанавливаются правилами внутреннего трудового распорядка или по соглашению между работником и работодателем.

Всем работникам предоставляются выходные дни (еженедельный непрерывный отдых). При пятидневной рабочей неделе работникам предоставляются два выходных дня в неделю, при шестидневной рабочей неделе - один выходной день.

Законодательством РФ запрещена дискриминация по любым признакам и принудительный труд.

Заключение

В данном отчете была рассмотрена программная платформа, представляющая собой комплекс моделей, алгоритмов и программ интеллектуального анализа данных, реализующий подход мета-обучения и основанный на методе группового учета аргументов.

Также были изучены подходы к прогнозированию временных рядов, теоретические основы Data mining. Был рассмотрен алгоритм МГУА и его применения для прогнозирования.

Для улучшения репрезентативности программной платформы был проведен поиск, изучение и анализ мета-характеристик, не присутствующих в программной платформе, но являющихся перспективными для расширения платформы. В итоге был подобран ряд перспективных мета-характеристик, которые были реализованы для улучшения репрезентативности результатов работы программной платформы.

Также реализованные модули были протестированы на устойчивость к зашумленности временного ряда и показали хорошие результаты. К тому же, исходя из планируемого переноса платформы на Python, модули были перенесены на язык программирования Python.

Была проведена работа по оценке ресурсоэффективности, в результате чего был сделан вывод о высокой эффективности решения, в отличии от конкурирующих.

Был проведен анализ решения по социальной ответственности, в результате чего были вынесены правила применения решения на предприятии и подчеркнуто, что решение не несет собой никакой опасности для экологии.

Список использованных источников

1. An Introductory Study on Time Series Modeling and Forecasting: Basic Concepts of Time Series Modeling / Ratnadip Adhikari, R. K. Agrawal – P. 12-29.
2. Time-series forecasting / Chris Chatfield – P. 10-55.
3. Data Mining: Concepts and Techniques. Third Edition / Jiawei Han, Micheline Kamber, Jian Pe - P. 5-32.
4. Модели, алгоритмы и программная платформа для реализации мета-обучения на основе метода группового учета аргументов / Орлов А.А. – С. 101-164
5. Datasets meta-features description for recommending feature selection algorithm / Andrey Filchenkov, Arseniy Pendryak – P. 13-17
6. The development of self-organization techniques in modelling: a review of the group method of data handling (GMDH) / L. Anastasakis, N. Mort – P. 2-27.
7. Нейросетевые методы обработки хаотических процессов / В. А. Головкин – С. 45-87.
8. Методы Ляпунова и их применение. / Зубов В. И. - Изд-во ЛГУ, 1957.
9. Hurst exponent and financial market predictability / Bo Qian, Khaled Rasheed – P. 2-8.
10. Показатель Хёрста и его скрытые свойства // Ю. А. Калущ, В. М. Логинов – Сибирский журнал индустриальной математики 2002. — С. 29-37.
11. Singular Spectrum Analysis for Time Series / Nina E. Golyandina, Anatoly Zhigljavsky – P. 11-30.
12. Метод «Гусеница» – SSA: Анализ временных рядов / Н. Э. Голяндина – С. 5-54.
13. Hurst parameter estimate [Электронный ресурс] / Chu Chen - URL: <https://www.mathworks.com/matlabcentral/fileexchange/19148-hurst-parameter-estimate>, свободный – Дата обращения: 06.06.2018

14. Lyapunov exponent estimation from a time series. [Электронный ресурс] / Alan Wolf – URL: <https://www.mathworks.com/matlabcentral/fileexchange/48084-lyapunov-exponent-estimation-from-a-time-series-documentation-added>, свободный - Дата обращения: 06.06.2018
15. Nonlinear measures for dynamical systems (based on one-dimensional time series) [Электронный ресурс] / Christopher Schölzel - URL: <https://pypi.python.org/pypi/nolds/0.4.1>, свободный - Дата обращения: 06.06.2018
16. Парадигма программирования: курс лекций / Л. В. Городняя ; Новосиб. гос. ун-т. – Новосибирск : РИЦ НГУ, 2015. – 206 с.
17. Основы функционального программирования / Л. В. Городняя – С. 3-60.
18. Динамический хаос (курс лекций): Лекция 3. Система Лоренца; Лекция 4. Динамика системы Лоренца. / Кузнецов С. П. - М.: Физматлит, 2001.
19. ГОСТ 12.0.003-74. ССБТ. Опасные и вредные производственные факторы. Классификация.
20. ГОСТ 12.1.045–84 ССБТ. Электростатические поля. Допустимые уровни на рабочих местах и требования к проведению контроля.
21. ГОСТ 12.1.038–82 ССБТ. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов.
22. ГОСТ 12.1.010-76 ССБТ. Взрывобезопасность. Общие требования.
23. ГОСТ 12.1.004-91 ССБТ. Пожарная безопасность. Общие требования.
24. СанПиН 2.2.2/2.4.1340–03. Санитарно-эпидемиологические правила и нормативы «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы».
25. Приказ Минздравсоцразвития РФ от 13.08.2009 N 588н.

Приложение А

Разделы на иностранном языке

Раздел 1

SUBJECT AREA

Студент:

Группа	ФИО	Подпись	Дата
8ИМ6А	Гайнудинов Тимур Сохибназарович		

Научный руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Буркатовская Ю.Б.	к.ф. – м.н.		

Консультант – лингвист отделения иностранных языков ШБИП:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИЯ	О.В.Комиссарова	к.ф.н.		

Консультант отделения информационных технологий ИШИТР:

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

1. Subject area

1.1. Time series

The time series is a set of statistics that maps the process under study and is time-bound. Each element of this data is called a measurement, or a level relative to a specific point in time. On this basis, the order of the elements in the time series is important. This is the essential difference between the time series and the simple sampling of data.

Since time series are a set of two elements, namely:

- numeric value;
- a timestamp that corresponds to a numerical value

time series are of different types:

- one-dimensional and multidimensional time series, differing in the number of indicators at one point in time;
- moment and interval time series that differ in numerical value at a certain point in time. In time series, a numerical value characterizes the process at a certain point in time, and in interval time series - for a certain period of time;
- equidistant and non-equidistant time series, characterized by the type of intervals between time stamps; intervals can be equal or different from each other;
- complete and incomplete time series, ie, with the absence or presence of missing values.

The analysis of time series is a set of various statistical and mathematical methods used to identify statistical characteristics, the structure of the time series and some other indicators. Statistical characteristics give an idea of the properties of the time series. The concept of the time series structure is important for constructing a mathematical model of the process under study, which is the source of the time series. Time series analysis also includes the task of identifying time series

components such as trend, cyclical, seasonal components and noise, which is residual information that can not be explained.

All of the above data are used in the tasks of predicting time series that arise in the problem of effective decision-making. The forecast is the construction of a time series model, based on the history of the process being studied to obtain the probable value of future data. The task of forecasting time series is one of the most important tasks when working with time series. This task occurs in almost all areas of human life, for example, weather forecasting, forecasting the value of securities, etc. [1]

1.2.Forecasting time series

To date, there is a huge number of models for predicting time series. Among them there are both simple (for example, linear regression) and complex (model based on transfer functions). They can be classified into several groups:

- regression models - models based on the search for a relationship between one variable and a variety of other factors
 - simple linear regression
 - multiple regression
 - nonlinear regression
- autoregressive models - a model based on the assumption that the following process states depend on several previous states of the system
 - ARIMAX - autoregression integrated moving average
 - GARCH - generalized autoregressive conditional heteroskedasticity
 - ARDLM - autoregression distributed lag model
- models of exponential smoothing - models based on smoothing the time series with the help of a weighted moving average

- exponential smoothing
- Holt model
- Holt-Winters model
- model based on the maximum likelihood sample - a model based on the idea of process repeatability, but with changed properties
- model on neural networks - a model that is a neural network
- model on Markov chains - a model based on the assumption that the next state of the process depends only on the current state and is in no way related to the previous ones
- models on classification-regression trees-a model based on the construction of classification trees, regression trees or classification-regression trees
- model based on the genetic algorithm - a model based on the construction of a genetic algorithm
- model on support vectors - a model based on the idea of transferring a time series to a space whose dimension exceeds the original one, and searching for a separating hyperplane with the maximum gap in this space

1.3. Data mining

In the modern world, the volumes of data collected and analyzed are simply enormous. A person is not able to analyze this amount of data independently, but the task of analyzing these data requires a solution, because these data hide useful knowledge, patterns, and dependencies that can be used to make decisions. Automatic data analysis is called intelligent data analysis, or Data mining [2].

Data mining is the process of discovering non-trivial useful information in large amounts of data that are not explicitly presented, but useful in a practical sense.

The obtained patterns and trends cannot be detected with traditional data viewing due to the complexity of the links and the excessive amount of data.

The main areas in which intelligent data analysis is used to solve problems:

Teaching with the teacher - this method involves obtaining a pair of values at the input: an object and the corresponding response. The task is to find and derive a functional relationship between the object and the response. The tasks solved with the help of training with the teacher are divided into:

- classification - the division of a given set of data into groups according to predetermined groups
- regression - the identification of patterns between certain characteristics or data
- ranking - obtaining results for several objects at once with subsequent sorting
- forecasting - obtaining future states of the object under study

Learning without a teacher - this method based on the received objects, in the absence of appropriate answers to these objects, should establish a connection between the objects. The tasks solved with the help of teaching without a teacher are divided into:

- clustering - grouping objects based on their characteristics, properties, parameters. An important condition is that objects within clusters should be similar, and should differ from objects from other clusters
- association - the identification of patterns between data
- Emission filtering - Detection of atypical objects in the general sample
- constructing a confidence area - selecting a region containing the desired sample fraction, minimizing the scope of the domain
- Dimension reduction - the transition from one feature space to another, smaller in dimension, while losing as little information as possible
- filling in the missing values - replacing missing data by building a forecast

Partial training is an intermediate link between teaching with a teacher and teaching without a teacher.

Transductive learning is the method that solves the problems of predicting data on one private sample for another private sample.

Training with reinforcement is a method that operates with data such as "situation" and "decision taken". The answer to the method is the function that determines the "correctness" of the decision.

Dynamic learning - the difference from other learning methods is that the data is flowing, which forces you to make a decision and to learn the model on new data at the same time.

Active learning is different as the trainee has the opportunity to independently designate the following precedent, which will become known.

Meta-learning is a method in which the solved learning tasks are a precedent. The tasks solved by meta-training are divided into:

- multitasking - a set of simultaneously solvable learning tasks
- Inductive transfer - the experience of solving individual particular learning tasks according to precedents is transferred to the solution of subsequent private learning tasks

2. Overview of the software platform

2.1. General information

The software platform under investigation was fully developed and implemented by the candidate of technical sciences Andrey A. Orlov. In the long term, the platform is not only of scientific interest, as is the automatic inductive method of forecasting, but also industrial, for example, for automatic forecasting of data in various fields or as an alternative to expert judgment.

This platform is a software package for implementing a meta-training system using the method of group accounting of arguments. The platform is a four-level system. At the last level, the system builds a forecast based on the model parameters and coefficients obtained at the higher level and adds the results to the database. At the third level, the coefficients for the selected model are calculated based on the best decision algorithm and model parameters. At the second level, the meta-characteristics of the time series are distinguished and, based on the information from the database, at the level above, the best algorithm and parameters for predicting the analyzed time series are selected. At the first level there is a database containing the results for each algorithm, with all possible parameters, up to a given limit, and a prediction error for each method. As a result, the system issues the 8 best algorithms for predicting the time series being analyzed, judging by the efficiency criterion. A detailed description can be found on the link [3].

3. Base of meta-characteristics

At the moment, the system presents 8 meta-characteristics of the time series:

1. **Length** - The length of the time sequence (the number of points).
2. **Noise** - The average value of the absolute change per step.

$$F_{\text{noise}} = \frac{1}{T} \sum_{t=1}^{T-1} |x_{t-1} - x_t|$$

3. **Trend** - the slope coefficient of a straight linear trend.

$$F_{\text{trend}} = a, x_t = a * t + b$$

4. **Period** –the period of the harmonic signal, calculated according to the ordinate balance criterion from the assumption of a single harmonic in the input time sequence.

A predictive model was constructed using the combinatorial algorithm of the GMDH method with the restriction to the first (land1, linear model) and the 7th (land2, nonlinear model) degree of the polynomial. The training sample consisted of the first half of the points in the time sequence, the test sample was the second. As an external criterion, the regularity criterion CR was used.

5. **Land1in** - The CR value in the training data sample

6. **Land1out** - The CR value in the data sampling sample

7. **Land2in** - The CR value in the training data sample

8. **Land2out** - The CR value in the data sampling sample

4. Statement of the task of adding meta-characteristics

The task of searching for new meta-characteristics arose during the testing of existing software. As a result of the tests, it became clear that when calculating meta-characteristics for different time series, important information about some properties of the time series is lost, namely, about noise and period.

On this basis, the task of improving the adequacy of the results of the method was to analyze the problem of information loss during the calculation of meta-characteristics and search for such meta-characteristics that will help to better track the meta-knowledge of the noise and periodic components of the time series being analyzed.

In the course of the search, a number of characteristics [4] were revealed, proposed in various sources and articles:

- Number of instances
- Number of features
- Number of classes

- Dataset dimensionality
- Standard deviation
- Variation coefficient
- Correlation coefficient
- Skewness
- Kurtosis
- Average normalized features entropy
- Normalized class attribute entropy
- Maximal mutual information between the attribute and the class
- Average mutual information between the attribute and the class
- Ratio of signal noise
- Number of equivalent features
- Weight sum on the full dataset
- Minimum weight sum
- Average weight sum
- Standard deviation of weight sum
- Number of neighbors for the full dataset
- Minimum number of neighbors
- Maximum number of neighbors
- Average number of neighbors
- Standard deviation of the number of neighbors

Criteria for selecting new meta-characteristics were:

- the data obtained as a result of calculating the meta-characteristics must be nontrivial and early unknown

- the data obtained as a result of calculating the meta-characteristics should contain information about the noise or periodic component of the analyzed process

The listed meta-characteristics, due to non-compliance with the criteria put up for the search for new meta-characteristics, were rejected.

Based on the results of the analysis, the following meta-characteristics were proposed:

Chaos - Chaos is a chaotic component, calculated with the help of the Lyapunov exponent.

Dynamic chaos is a phenomenon in the theory of dynamical systems, in which the behavior of a nonlinear system looks random despite the fact that it is determined by deterministic laws.

Noise Influence - this characteristic is calculated using spectral singular analysis (SSA).

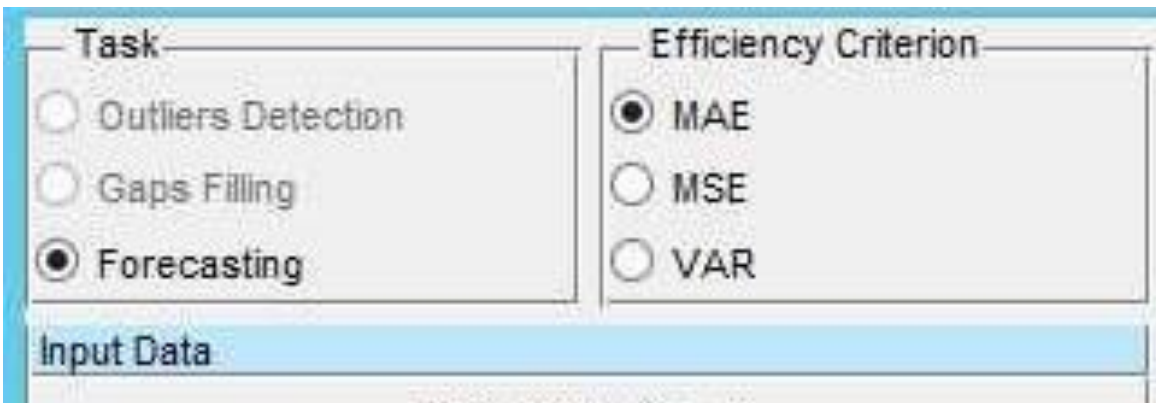
The detection of the influence of noise is carried out by calculating the difference of the noise space from the estimated signal subspace. Accordingly, if the difference begins to increase, this indicates a strong noisiness.

Discontinuity - The spasmodicity is determined by the indicator of Hurst.

One of the most important tasks of monitoring a technical object is the task of timely detection of significant changes in physical quantities (state variables) characterizing the state of the object. An abrupt change of even one variable, as a rule, signals the possibility of a serious transformation of the state of the object, which can lead to an emergency situation or an accident.

Приложение Б

В приложении Б представлены скриншоты работы исследуемой программной платформы.



The screenshot displays a user interface with two main sections for configuration. The left section, titled "Task", contains three radio button options: "Outliers Detection", "Gaps Filling", and "Forecasting". The "Forecasting" option is selected, indicated by a filled circle. The right section, titled "Efficiency Criterion", contains three radio button options: "MAE", "MSE", and "VAR". The "MAE" option is selected, indicated by a filled circle. Below these sections is a blue header for "Input Data".

Выбор задачи и критерия эффективности

Input Data

M3C_M3Month.csv

Category	N	ID
MICRO		68 1402
MICRO		68 1403
MICRO		68 1404
MICRO		68 1405
MICRO		68 1406
MICRO		68 1407
MICRO		68 1408
MICRO		68 1409
MICRO		68 1410
MICRO		68 1411
MICRO		68 1412
MICRO		68 1413
MICRO		68 1414
MICRO		68 1415
MICRO		68 1416
MICRO		68 1417
MICRO		68 1418
MICRO		68 1419
MICRO		69 1420
MICRO		69 1421
MICRO		69 1422
MICRO		69 1423
MICRO		69 1424
MICRO		69 1425
MICRO		69 1426
MICRO		69 1427
MICRO		69 1428
MICRO		69 1429
MICRO		69 1430
MICRO		69 1431
MICRO		69 1432
MICRO		69 1433
MICRO		69 1434
MICRO		69 1435
MICRO		69 1436
MICRO		69 1437
MICRO		69 1438
MICRO		69 1439
MICRO		69 1440
MICRO		69 1441
MICRO		69 1442
MICRO		69 1443
MICRO		69 1444
MICRO		69 1445
MICRO		69 1446
MICRO		69 1447
MICRO		69 1448
MICRO		69 1449

Refresh Data

Набор временных рядов, взятых из различных областей и используемых для анализа

Meta-Base_Cut.csv				
Load Meta-Base				
Task	MetaFeatures	Algorithm	Parameters	Efficien
Forecasting	length = 50; ...	MultilayerPer...	template = 1;...	CR
Forecasting	length = 50; ...	MultilayerPer...	template = 1;...	MAE
Forecasting	length = 50; ...	MultilayerPer...	template = 1;...	MSE
Forecasting	length = 50; ...	MultilayerPer...	template = 1;...	VAR
Forecasting	length = 50; ...	GaussianPro...	template = 1;...	CR
Forecasting	length = 50; ...	GaussianPro...	template = 1;...	MAE
Forecasting	length = 50; ...	GaussianPro...	template = 1;...	MSE
Forecasting	length = 50; ...	GaussianPro...	template = 1;...	VAR
Forecasting	length = 50; ...	LinReg	template = 1;...	CR
Forecasting	length = 50; ...	LinReg	template = 1;...	MAE
Forecasting	length = 50; ...	LinReg	template = 1;...	MSE
Forecasting	length = 50; ...	LinReg	template = 1;...	VAR
Forecasting	length = 50; ...	LinReg	template = 1;...	CR
Forecasting	length = 50; ...	LinReg	template = 1;...	MAE
Forecasting	length = 50; ...	LinReg	template = 1;...	MSE
Forecasting	length = 50; ...	LinReg	template = 1;...	VAR
Forecasting	length = 50; ...	LinReg	template = 1;...	CR
Forecasting	length = 50; ...	LinReg	template = 1;...	MAE
Forecasting	length = 50; ...	LinReg	template = 1;...	MSE
Forecasting	length = 50; ...	LinReg	template = 1;...	VAR
Forecasting	length = 50; ...	LinReg	template = 1;...	CR
Forecasting	length = 50; ...	LinReg	template = 1;...	MAE
Forecasting	length = 50; ...	LinReg	template = 1;...	MSE
Forecasting	length = 50; ...	LinReg	template = 1;...	VAR
Forecasting	length = 50; ...	LinReg	template = 1;...	CR

Update Meta-Base

Meta-models

Build Meta-models

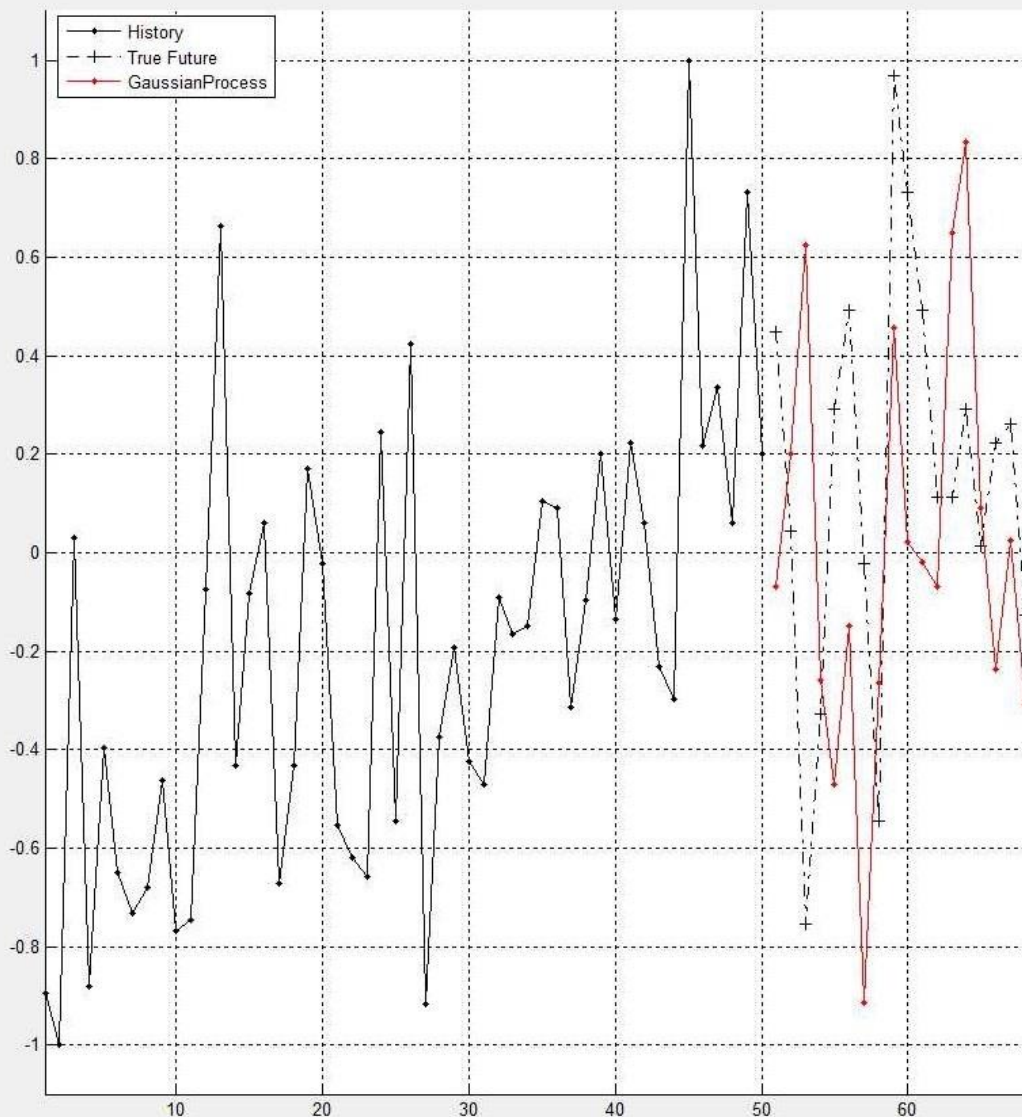
База данных, в которой содержатся результаты для каждого алгоритма, со всеми возможными, до заданного ограничения, параметрами, а также ошибка прогнозирования для каждого метода.

Meta-Features	
Name	Value
length	50
noise	0.4168
trend	0.9096
period	4.5995
land1in	0.1476
land1out	0.1186
land2in	0.1476
land2out	0.1186

Мета-характеристики анализируемого временного ряда

Algorithms			
Algorithm	Optimal Params	Efficiency Prediction	NMBRI
COMBI	max_pow = 1;	5.5845	21
COMBIHARM...	m_max = 1; max_pow = 1;	5.9788	21
GaussianPro...	gp_noise = 0.01; template = 4;	6.4712	24
HARMONIC	m_max = 4; mean = 1;	10.2866	14
LinReg	EliminateCollinear = 0; Ridge = 1e-08; SelectionMethod = M5; template ...	6.6282	61
MultilayerPer...	epochs = 500; hiddenLayers = a; learningRate = 0.3; momentum = 0.2...	5.2617	23
PNN	Z = 2; delta = 1e-05; max_pow = 2; neurons_n = 2; r_max = 2; templa...	5.1747	41
POLYHARM...	m_max = 1; max_pow = 1; mean = 1;	4.2312	21

Алгоритмы, проявившие себя наилучшим образом в процессе прогнозирования анализируемого временного ряда



Пример отображение реальных и прогнозируемых данных