

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
Кафедра компьютерных технологий и систем**

КОНДРАТОВИЧ Артём Игоревич

**ИНСТРУМЕНТЫ И РЕАЛИЗАЦИИ СРЕДСТВ
ИЗВЛЕЧЕНИЯ, ОБРАБОТКИ И ВИЗУАЛИЗАЦИИ
ЭКОНОМИЧЕСКОЙ ИНФОРМАЦИИ**

Дипломная работа

Научный руководитель:
профессор кафедры КТС,
доктор физ.-мат. наук,
профессор Таранчук В. Б.

Допущена к защите

« ____ » _____ 20 ____ г.

Заведующий кафедрой компьютерных технологий и систем
доктор педагогических наук, кандидат физико-
математических наук, профессор В.В. Казачёнок

Минск, 2025

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	7
Глава 1 ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ВОПРОСЫ	9
1.1 Общие сведения об экономической информации	9
1.2 Классификация экономического анализа	13
Глава 2 ТИПЫ (КАТЕГОРИИ) ГРАФИЧЕСКИХ ОБЪЕКТОВ WOLFRAM MATHEMATICA.....	17
2.1 Примеры графических объектов	17
2.2 Финансовые индикаторы.....	21
Глава 3 ОБЗОР БИБЛИОТЕК ДЛЯ ВИЗУАЛИЗАЦИИ И АНАЛИЗА ДАННЫХ НА ПЛАТФОРМЕ .NET	25
3.1 Общие сведения о платформе .NET	25
3.2 Библиотека ScottPlot	26
3.3 Библиотека Skender.Stock.Indicators.....	28
Глава 4 ПРОГНОЗИРОВАНИЕ ФИНАНСОВЫХ ДАННЫХ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ	30
4.1 Линейная регрессия	30
4.2 Дерево решений.....	32
4.3 Методы ансамблирования	32
4.4 Сравнительный анализ методов	33
Глава 5 СИСТЕМА DIRECTUM RX. ИНТЕГРАЦИЯ С АВТОМАТИЗИРОВАННОЙ ИНФОРМАЦИОННОЙ СИСТЕМОЙ «ВЕДОМСТВЕННЫЙ АРХИВ» (АИС ВА).....	36
5.1 Системы электронного документооборота	36
5.2 Система Directum RX.....	37
5.3 Автоматизированная информационная система «Ведомственный архив» (АИС ВА)	38
5.4 Интеграция с АИС ВА	40
Глава 6 ПРИМЕРЫ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ	49
6.1 Извлечение и подготовка данных.....	49
6.2 Визуализация полученных данных	50
6.3 Расчёты прогнозов	52

6.4 Сравнение прогнозов с реальными данными	54
6.5 Реализация на платформе .NET	55
6.6 Сравнительный анализ полученных результатов	58
ЗАКЛЮЧЕНИЕ	59
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	60

РЕФЕРАТ

Структура и объём дипломной работы

60 страниц, 52 рисунка, 1 таблица, 9 источников

Ключевые слова: WOLFRAM MATHEMATICA, ИНФОРМАЦИЯ, ДАННЫЕ, ВИЗУАЛИЗАЦИЯ, АНАЛИЗ, ПРОГНОЗИРОВАНИЕ, СИСТЕМЫ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА (СЭД), ИНТЕГРАЦИЯ.

Текст реферата

Объект исследования – экономическая информация, её разновидности и способы её извлечения, визуализации и интеллектуальной обработки.

Цель исследования – всестороннее изучение возможностей системы Wolfram Mathematica, как инструмента для полного цикла работы с экономической информацией и сравнительный анализ с другими реализациями.

Методы исследования – изучение соответствующей литературы и электронных источников, программирование на языках Wolfram Language и C# и изучение существующих подходов к работе с экономической информацией.

Полученные результаты – описанные возможности системы Wolfram Mathematica для решения задач извлечения, обработки, визуализации и прогнозирования экономических данных. Также проделан сравнительный анализ точности прогнозирования на разных платформах, на основании которого сделан вывод о том, что система Wolfram Mathematica обладает более точными реализациями методов прогнозирования. Также в рамках преддипломной практики реализована интеграция СЭД “Directum RX” с АИС «Ведомственный архив».

Область возможного применения – полученная информация в рамках исследования системы Wolfram Mathematica, как полноценного инструмента для работы с экономической информацией, может использоваться как рекомендательный материал для пользователей выбирающих инструменты для работы с экономической информацией. А имеющиеся в работе разобранные примеры – использовать в качестве обучающего материала. Что же касается результатов преддипломной практики, то область применения данной интеграции достаточно широка и подойдёт абсолютно каждой организации, которая использует систему “Directum RX” и заинтересована в цифровизации архивного хранилища документов.

РЭФЕРАТ

Структура і аб'ём дыпломнай працы

60 старонак, 52 малюнка, 1 табліца, 9 крыніц

Ключавыя словы: WOLFRAM MATHEMATICA, ІНФАРМАЦЫЯ, ДАДЗЕННЫЯ, візуалізацыя, АНАЛІЗ, ПРАГНАЗАВАННЕ, СІСТЭМЫ ЭЛЕКТРОННАГА ДАКУМЕНТААБАЛЁТУ (СЭД), ІНТЭГРАЦЫЯ.

Змест працы

Аб'ект даследавання – эканамічная інфармацыя, яе разнавіднасці і спосабы яе вымання, візуалізацыі і інтэлектуальнай апрацоўкі.

Мэта даследавання – ўсебаковае вывучэнне магчымасцей сістэмы Wolfram Mathematica, як інструмента для поўнага цыкла працы з эканамічнай інфармацыяй і параўнальны аналіз з іншымі рэалізацыямі.

Метады даследавання – вывучэнне адпаведнай літаратуры і электронных крыніц, праграмаванне на мовах Wolfram Language і C# і вывучэнне існуючых падыходаў да працы з эканамічнай інфармацыяй.

Атрыманыя вынікі – апісаныя магчымасці сістэмы Wolfram Mathematica для рашэння задач вымання, апрацоўкі, візуалізацыі і прагназавання эканамічных дадзеных. Таксама праведзены параўнальны аналіз дакладнасці прагназавання на розных платформах, на падставе якога зроблены выснова аб тым, што сістэма Wolfram Mathematica валодае больш дакладнымі рэалізацыямі метадаў прагназавання. Таксама ў рамках пераддыпломнай практыкі рэалізавана інтэграцыя СЭД "Directum RX" з АІС "Ведамасны архіў".

Вобласць магчымага практычнага прымянення – атрыманая інфармацыя ў рамках даследавання сістэмы Wolfram Mathematica, як паўнаўвартаснай прылады для працы з эканамічнай інфармацыяй, можа выкарыстоўвацца як рэкамендацыйны матэрыял для карыстачоў абіральных прылады для працы з эканамічнай інфармацыяй. А наяўныя ў працы разабраныя прыклады – выкарыстоўваць у якасці навучальнага матэрыялу. Што ж тычыцца вынікаў пераддыпломнай практыкі, то вобласць прымянення дадзенай інтэграцыі дастаткова шырокая і падыходзіць абсалютна кожнай арганізацыі, якая выкарыстоўвае сістэму "Directum RX" і зацікаўлена ў цыфравізацыі архіўнага сховішча дакументаў.

SUMMARY

Structure and scope of the diploma work

60 pages, 52 figures, 1 table, 9 references

Keywords: WOLFRAM MATHEMATICA, INFORMATION, DATA, VISUALIZATION, ANALYSIS, FORECASTING, ELECTRONIC DOCUMENT MANAGEMENT SYSTEMS (EDMS), INTEGRATION.

Summary text

The object of the research – economic information, its types and methods of its extraction, visualization and intellectual processing.

The aim of the research – a comprehensive study of the capabilities of the Wolfram Mathematica system as a tool for a full cycle of work with economic information and a comparative analysis with other implementations.

Research methods – study of relevant literature and electronic sources, programming in the Wolfram Language and C#, and study of existing approaches to working with economic information.

The results of the work – described the capabilities of the Wolfram Mathematica system for solving problems of extracting, processing, visualizing and forecasting economic data. A comparative analysis of the accuracy of forecasting on different platforms was also carried out, on the basis of which it was concluded that the Wolfram Mathematica system has more accurate implementations of forecasting methods. Also, as part of the pre-graduation practice, the integration of the Directum RX EDMS with the Departmental Archive AIS was implemented.

Recommendations on the usage – the information obtained in the study of the Wolfram Mathematica system as a full-fledged tool for working with economic information can be used as a recommendation for users choosing tools for working with economic information. And the analyzed examples available in the work can be used as training material. As for the results of pre-graduation practice, the scope of application of this integration is quite wide and will suit absolutely every organization that uses the Directum RX system and is interested in digitalizing the archival storage of documents.

ВВЕДЕНИЕ

В условиях стремительной цифровизации экономики и большого роста объема данных проблема эффективной обработки, анализа и визуализации экономической информации актуальна как никогда. Если разобрать структуру экономических данных, то они представляют собой сложный комплекс взаимосвязанных показателей, таких как отчеты, финансовые результаты и рыночные тренды. Ввиду этой сложности структуры данных формируется основная задача визуализации и анализа экономической информации – выявление ключевых показателей и трендов, которые помогают оценить состояние экономики, определить сильные и слабые стороны бизнеса, выявить потенциальные возможности для роста и развития. Благодаря визуализации данных становится возможным обнаружить скрытые зависимости, причинно-следственные связи и факторы, влияющие на экономические процессы. В данном контексте методы визуализации и анализа данных становятся критически важными инструментами для принятия обоснованных управленческих решений.

Степень разработанности данной проблемы в научной литературе остается достаточно неоднозначной: несмотря на обилие исследований, посвященных методам анализа данных (включая машинное обучение, статистические модели), вопросы, связанные с использованием универсальных вычислительных систем, таких как Wolfram Mathematica, в экономическом анализе изучены недостаточно. Большая часть существующих работ ориентированы на узкоспециализированные системы, тогда как система Wolfram Mathematica в области экономических исследований требует дополнительного изучения.

Теоретическая значимость проделанной работы заключается в систематизации подходов к интеллектуальному анализу и визуализации экономических данных с использованием Wolfram Mathematica.

Практическая ценность исследования состоит в оценке эффективности Wolfram Mathematica по сравнению с альтернативными решениями, что может послужить основанием для рекомендаций по его применению для полного цикла обработки экономических данных.

Целью дипломной работы является всесторонний анализ системы Wolfram Mathematica, как универсального инструмента для работы с экономической информацией. В моей работе поставлены следующие цели:

- Исследовать функциональные возможности Wolfram Mathematica для сбора, обработки визуализации и прогнозирования экономических данных.

- Провести сравнительный анализ с альтернативными программными решениями.
- Оценить эффективность прогнозирования данных с помощью рассмотренных систем и провести сравнительный анализ с целью выявления более точной системы.

Глава 1

ТЕОРЕТИКО-МЕТОДОЛОГИЧЕСКИЕ ВОПРОСЫ

1.1 Общие сведения об экономической информации

Первоначально смысл слова «информация» (от лат. «*informātiō*») трактовался как нечто присущее только человеческому сознанию и общению - знания, сведения, известия. Затем смысл этого слова начал расширяться и обобщаться. Информация содержится в человеческой речи, текстах книг, журналов и газет, сообщениях радио и телевидения, показаниях приборов и т. д. Человек воспринимает информацию с помощью органов чувств, хранит и перерабатывает ее с помощью мозга и центральной нервной системы. Передаваемая информация обычно касается каких-то предметов или нас самих и связана с событиями, происходящими в окружающем нас мире.

Любая информация обладает рядом свойств и характеристик, которые в совокупности определяют степень ее соответствия потребностям пользователя (качество информации). Можно привести немало разнообразных свойств и характеристик информации, так как каждая научная дисциплина рассматривает только те, которые ей наиболее важны. С точки зрения экономики наиболее важными представляются следующие характеристики:

- Корректность
- Ценность
- Достоверность
- Точность
- Актуальность
- Полнота

и свойства:

- большие объемы обработки информации;
- многократное повторение ее циклов получения и преобразования в установленные периоды;
- многообразие источников и потребностей ее информации;
- значительный удельный вес арифметических операций;
- точность экономической информации;
- достоверность;
- оперативность (актуальность).

Экономическая информация – это информация о процессах производства, и потребления материальных благ и различных услуг. Она представляет собой числовые данные, отображающие экономическую деятельность любого предприятия. Пример данных на рисунках 1.1-1.3.

С течением времени объем экономической информации, вокруг нас, стал стремительно расти и параллельно с ростом усложняется и понимание экономической информации, поэтому требуется как-то анализировать данные и представлять их в удобном для работы формате, поэтому анализ и визуализация экономической информации весьма актуальны в наше время.

Анализ – это разложение сложного на составные части, с целью понимания исследуемого объекта, явления.

Экономический анализ – одна из разновидностей анализа, представляющий собой аналитическую процедуру, целью которой является – вывод, заключение экономического характера, в отношении некоторого результата деятельности организации.

Основная цель экономического анализа – помощь в принятии правильных и своевременных управленческих решений на некотором временном интервале.

Каким должен быть экономический анализ? Он должен быть:

- конкретным;
- целенаправленным;
- действенным;
- оперативным;
- систематическим;
- достоверным;
- объективным



Рисунок 1.1 – Пример экономических данных компании ЕРАМ за март 2023



Рисунок 1.2 – Пример экономических данных компании ЕРАМ за март 2024

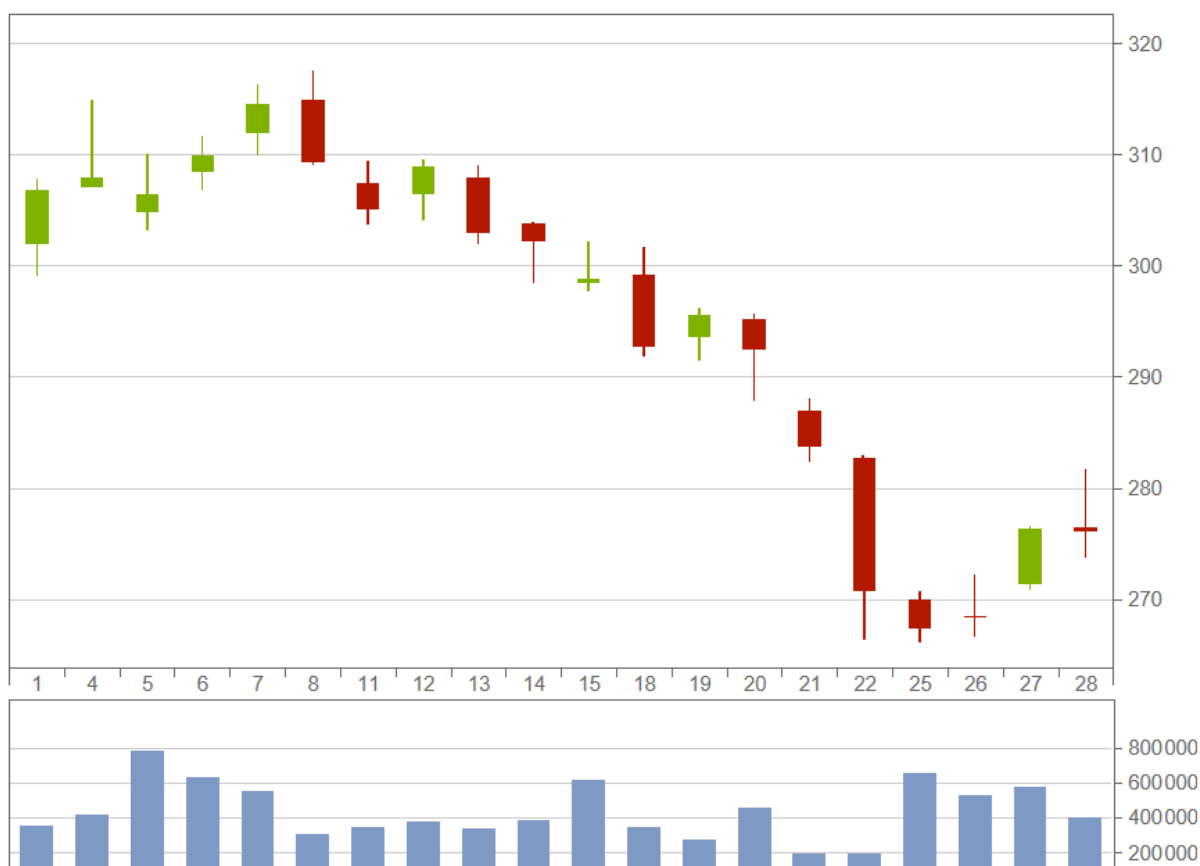


Рисунок 1.3 – Пример экономических данных компании ЕРАМ за март 2025

Важной характеристикой данных является её представительность. Под представительностью я подразумеваю оценку качества предварительной обработки и визуализации текстовой и числовой информации (самих данных) и чем качественнее предварительная обработка и грамотнее подобраны способы графического представления данных, тем выше их представительность. Высокая представительность данных позволяет целевой аудитории анализировать текущее состояние, прогнозировать, принимать решения.

Рассмотрим основные принципы представительности данных. Первым из них является целевая аудитория. В зависимости от целевой аудитории необходимо тщательно выбирать формат представления данных, уровень сложности и стиль и для достижения высокой представительности необходимо знать уровень понимания аудитории в рассматриваемой вами области и чем выше будет этот уровень, тем сложнее может быть структура представления данных. Это включает в себя использование специализированных терминов, сложных графиков и глубокого анализа, но это не отменяет того, что отображаемая информация ясна и доступна. Если мы говорим про более широкую публику, то естественно опускаются технические термины, сложные графики и делается упор на простоту визуализации и подачу информации.

Второй принцип актуальность и точность данных. По своей сути соблюдая данный принцип мы закладываем прочный фундамент для достижения высокого уровня представительности. Ведь без соблюдения данного принципа, все сделанные выводы, принятые решения и возможные прогнозы будут с большой вероятностью неправильными, что будет приводить к потере времени и ресурсов.

Также немало важно это умение рассказать о данных. Хотя тщательно обработанные данные с удачно подобранной визуализацией уже являются представительными, этот показатель всегда можно повысить за счет рассказа о данных, где рассказчик отмечает важность и актуальность данных в рассматриваемой предметной области, показывает логические связи, даёт обоснование неочевидному поведению данных и т.д. Всё это ускоряет процесс восприятия информации и помогает полностью разобраться в ней.

На сегодняшний день в областях, тесно связанных с данными, такая характеристика, как представительность, является одной из ключевых и рассмотренные выше рекомендации и принципы помогут эффективно использовать данные, что в свою очередь будет способствовать развитию.

1.2 Классификация экономического анализа

Так как экономический анализ это – вывод, заключение в отношении экономической деятельности организации, а разные организации имеют различные цели и задачи, поэтому требуются определённые виды экономического анализа, которые можно классифицировать [1,4]:

- по содержанию и полноте изучаемых объектов;
- по методам изучения объектов;
- по периодичности;
- по признаку времени;
- по объектам управления.

По содержанию и полноте изучаемых объектов, делится на:

- Комплексный анализ – всестороннее изучение экономической деятельности предприятия;
- Локальный анализ – изучение конкретного подразделения;
- Тематический анализ – изучение наиболее важных в определённое время вопросов.

Существуют следующие методы изучения:

- Сравнительный – чаще всего это сравнение отчётных показателей с показателями текущего плана и показателями компаний-конкурентов;
- Факторный анализ направлен на количественное описание влияния различных факторов на экономические показатели;
- Маржинальный анализ представляет собой метод, который используется для оценки и обоснования эффективности управленческих решений в бизнесе. Он основан на анализе причинно-следственной связи между объемом продаж, себестоимостью и прибылью, а также на разделении затрат на постоянные и переменные;
- Экономико-математический – это метод анализа, представляющий собой объединение экономической теории и математических моделей для описания, прогнозирования и оптимизации экономических систем;
- Стохастический является ветвью математики и статистики, чаще всего используется, когда требуется проанализировать экономическую систему, где преобладают неопределённость и случайность;
- Диагностический анализ применяется с целью анализа и понимания причин проблем и выявления неисправностей в системе предприятия или конкретного процесса.

Классификация экономического анализа по периодичности помогает организовать сбор данных и анализировать их в соответствии с требуемым временным интервалом. Различные периоды анализа предоставляют разную информацию и позволяют оценить экономическую динамику на разных уровнях детализации и временных горизонтах. Опираясь на это, можно выделить следующие типы:

- Годовой анализ проводится ежегодно. Позволяет оценить долгосрочные тенденции и изменения в экономике, такие как рост ВВП, инфляция, безработица, доходы и другие показатели. Он широко используется для оценки экономического развития страны или региона;
- Квартальный анализ сбор и анализ проводится каждый квартал (каждые три месяца). Отличается от остальных возможностью получения более оперативной информации о текущем состоянии экономики;
- Декадный анализ чаще всего используется в финансовой сфере для отслеживания краткосрочных изменений на рынке, таких как цены на товары, колебания курсов валют, ставки процента и другие финансовые показатели. Проводится данный анализ раз в 10 дней;

- Разовый анализ применяется для получения информации в конкретный момент для конкретной цели. Это может быть анализ финансовой отчётности компании за какой-то специфический период времени, который не удастся получить, используя вышеизложенные типы;
- Ежедневный анализ проводится каждый день и в силу такого малого временного интервала в основном он используется для анализа краткосрочных тенденций и изменений на рынке.

По признаку времени можно выделить два основных типа экономического анализа:

- Перспективный (прогнозный) анализ проводится на основе данных, относящейся к будущему периоду времени. Зачастую применяется предприятиями для прогнозирования результатов в будущем с целью планирования, принятия управленческих решений и определения стратегии развития;
- Ретроспективный (исторический) анализ используется для контроля выполнения программ развития предприятия, объективной оценки ранее принятых управленческих решений и выбранных стратегий развития. В зависимости от глубины и качества данного анализа зависят результаты следующего прогноза, а значит, данный анализ является основой для перспективного анализа. Ретроспективный анализ разделяется на оперативный и текущий. Основная задача оперативного анализа – это оперативное выявление недостатков, оперативность достигается за счёт того, что его проводят сразу после совершения хозяйственных актов. Если говорить о текущем анализе, то он проводится на основе периодического анализа, на основании которого и формируется объективная оценка деятельности.

В экономическом анализе объект управления относится к субъекту или сфере деятельности, на которые направлен анализ. Рассмотрим классификацию экономического анализа по объектам управления.

Технико-экономический анализ фокусируется на изучении взаимосвязи между техническими и экономическими параметрами предприятий. Включается в себя оценку эффективности использования ресурсов, расчёт затрат и доходов, анализ технологических процессов и определение оптимальных технических параметров производства. Одним из примеров может быть использование энергосберегающих технологий производства.

В финансово-экономическом анализе основное внимание уделяется финансовым результатам деятельности предприятия. Предметом анализа являются финансовые отчёты, такие как балансовые ведомости, отчеты о

прибылях и убытках, денежные потоки. На основе полученных данных формируется объективная оценка финансового состояния предприятия.

Экономико-статистический анализ применяется для изучения массовых общественных явлений на разных уровнях управления: предприятия, отрасли, региона.

Социально-экономический анализ оценивает социальные и экономические последствия различных действий и проектов. Он включает изучение влияния экономических решений на общество, социальные группы и отдельных людей, оценку социальных индикаторов (уровень бедности, занятость, социальная защищенность) и социальных последствий экономической деятельности.

Экономико-экологический анализ занимается изучением взаимосвязи между экономической деятельностью и ее воздействием на окружающую среду. Он направлен на оценку влияния экономических процессов на экологическое состояние и на определение оптимальных путей сохранения и улучшения окружающей среды с эффективным использованием ресурсов и снижением экологических затрат.

Маркетинговый анализ играет важную роль в ведении бизнеса, поскольку позволяет оценивать рыночный потенциал, конкурентную позицию и эффективность маркетинговых стратегий. Он помогает предприятиям адаптироваться к изменениям на рынке, выявлять новые возможности и принимать меры для удовлетворения потребностей клиентов, что способствует достижению конкурентных преимуществ и повышению результативности бизнеса.

Управленческий анализ затрагивает все сферы деятельности предприятия. В основном направлен на изучение и оценку эффективности и эффективности использования ресурсов предприятия, а также на принятие решений внутри организации.

Каждый из перечисленных видов экономического анализа различен по содержанию, организации и методике его проведения. В практической деятельности используются одновременно несколько видов анализа, представляющие единую систему принятия эффективных управленческих решений.

Глава 2

ТИПЫ (КАТЕГОРИИ) ГРАФИЧЕСКИХ ОБЪЕКТОВ WOLFRAM MATHEMATICA

Wolfram Mathematica – это одна из мощнейших систем компьютерной математики, которая обладает обширным набором функций для работы с данными. Такое количество встроенных функций и способов модификаций их с помощью различных опций, директив, атрибутов даёт возможность для создания различных графиков и разнообразие в визуализации данных.

Что же собой представляет графический объект в системе Wolfram Mathematica? Графический объект – это объект, возвращаемый графической функцией, с которым можно выполнять следующие операции:

- Визуализация. Представляет собой операцию отображения данного объекта на экране в некотором виде. Это может быть диаграмма, гистограмма, карта, график и т.д.;
- Манипуляция. Под данной операцией подразумевается любое изменение встроенных полей (свойств) графического объекта. Это может быть изменение цвета, размера, стиля, текста, подписи и многое другое;
- Композиция. Исходя из названия данная операция подразумевает композицию из нескольких графических объектов, с целью создания более сложных объектов под определённые задачи;
- Экспорт и публикация. Все графические объекты могут быть экспортированы в различных форматах файлов.

Для работы с экономической информацией, а именно анализ и визуализация, нам понадобится категория графических объектов, графики функций, графики данных. Рассмотрим некоторые примеры.

2.1 Примеры графических объектов

Графики аналитически задаваемых функций одной переменной [7,8]:

- Plot – график аналитической заданной функции $f(x)$ (и везде ниже – или нескольких функций), обе шкалы линейные;
- LogPlot – LogPlot[f, {x, x_{min} , x_{max} }] генерирует график аналитически заданной функции $f(x)$ от x_{min} до x_{max} в логарифмической шкале (по оси ординат шкала логарифмическая, по оси абсцисс - линейная);

- `LogLinearPlot` – `LogLinearPlot[f, {x,  $x_{min}$ ,  $x_{max}$ }]` генерирует график аналитически заданной функции $f(x)$ от x_{min} до x_{max} в логарифмически линейной шкале (по оси ординат шкала линейная, по оси абсцисс логарифмическая);
- `LogLogPlot` – `LogLogPlot[f, {x,  $x_{min}$ ,  $x_{max}$ }]` генерирует график аналитически заданной функции $f(x)$ от x_{min} до x_{max} в лог-лог шкале (логарифмический по обеим осям);
- `PolarPlot` – `PolarPlot[r, { $\theta$ ,  $\theta_{min}$ ,  $\theta_{max}$ }]` генерирует график кривой радиуса r в зависимости от θ угла в полярной системе координат - положение конца радиус-вектора при изменении угла в заданном диапазоне (полярная диаграмма);
- `ParametricPlot` – график кривой (а кривой (а также поверхности), заданной параметрически, `ParametricPlot[{ $f_x$ ,  $f_y$ }, { $u$ ,  $u_{min}$ ,  $u_{max}$ }]` генерирует график кривой с x и y координатами f_x и f_y как функций u . Рассмотрим его пример на рисунке 2.1.

```
f[u_] := {u * Cos[3 * u], u * Sin[2 * u]}
ParametricPlot[f[u], {u, 0, 5 * Pi}, ImageSize -> 270, Mesh -> 4,
  MeshFunctions -> {#3 &}, MeshShading -> {Orange, Purple}]
```

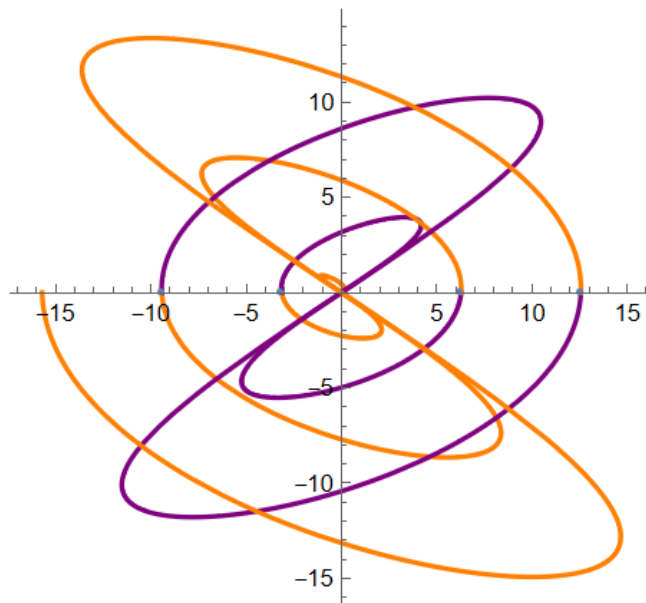


Рисунок 2.1 – Пример использования ParametricPlot

Визуализация массивов одномерных данных [1,4]:

- ListPlot – диаграмма разброса данных; ListPlot[{y1, y2, ...}] выводит график списка величин y1, y2, ..., координаты которых задаются значениями номеров 1, 2, ...;
- ListLinePlot – Линейный график по точкам списка данных, ListLinePlot[{y1, y2, ...}] выводит график, на котором соединяет отрезками последовательные точки с координатами (значение, номер);
- ListStepPlot – ступенчатый график по списку значений, ListStepPlot[{y1, y2, ...}]] строит график, на котором соединяет ступеньками последовательные точки с координатами (значение, номер);
- ListLogPlot – диаграмма разброса данных в лог-шкале (по оси ординат шкала логарифмическая, по оси абсцисс линейная);
- ListLogLinearPlot – диаграмма разброса данных в лог-линейной шкале (по оси ординат шкала линейная, по оси абсцисс логарифмическая);
- ListLogLogPlot – диаграмма разброса данных в лог-лог шкале (логарифмический по обеим осям);
- ListPolarPlot – диаграмма разброса данных в полярных координатах;

- `DiscretePlot` – генерирует график последовательности, `DiscretePlot[expr, {n, nmax}]` генерирует график определяемых по формуле *expr* значений *n* для последовательности от 1 до *n_{max}*;
- `NumberLinePlot` – диаграмма на числовой оси, `NumberLinePlot[{v1, v2, ...}]` выводит значения *v_i* на числовой линии;
- `DateListPlot[{{date1, v1}, {date2, v2}, ...}]` генерирует точки со значениями *v_i* на заданные даты;
- `DateListStepPlot[{{date1, v1}, {date2, v2}, ...}]` генерирует график ступеньками по точкам со значениями *v_i* и заданными датами;
- `DateListLogPlot[{{date1, v1}, {date2, v2}, ...}]` генерирует график по точкам со значениями *v_i* и заданными датами (по оси ординат шкала логарифмическая);
- `QuantilePlot` – квантильный Q-Q график (для сравнения совокупности данных с согласованным нормальным распределением, оценки качества подгонки к наблюдаемым данным);
- `ProbabilityPlot` – график плотности распределения, `ProbabilityPlot[list]` генерирует график CDF (cumulative distribution function – функция распределения) случайной величины *list*;
- `ProbabilityScalePlot[{x1, x2, ...}]` генерирует график плотности распределения (в процентах) по набору *x_i*. Рисунок 2.2.

```
data = {1.2, 1.5, 1.8, 2.1, 2.2, 2.6, 2.8, 2.9, 3.1, 3.5, 3.9, 4.2, 4.5, 4.7, 5.2};
ProbabilityScalePlot[data, PlotRange -> All, Frame -> True, FrameLabel -> {"Theoretical Quantiles", "Ordered Values"},
PlotLabel -> "Probability Scale Plot", PlotStyle -> Directive[PointSize[Medium], Red], ImageSize -> 600]
```

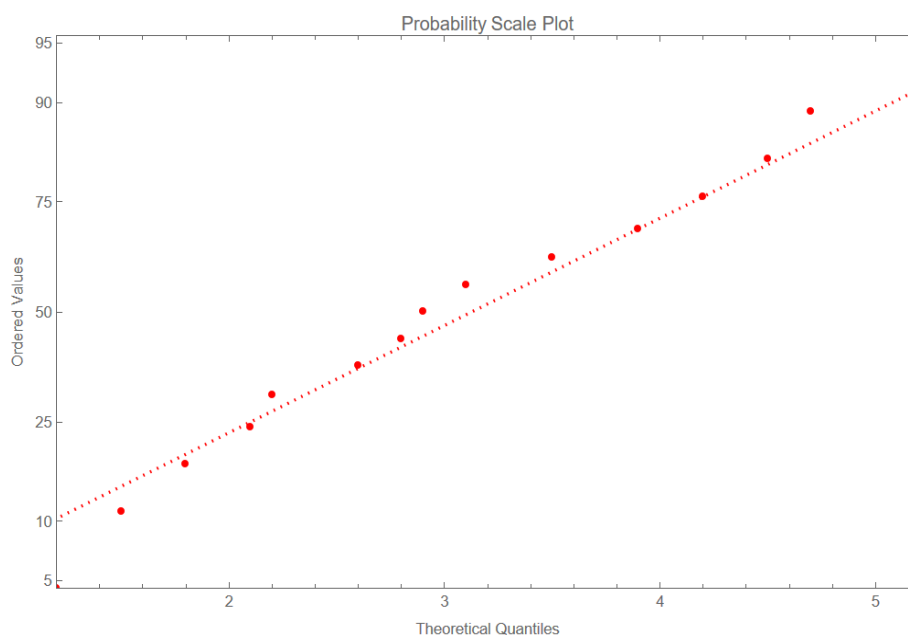


Рисунок 2.2 – Пример использования ProbabilityScalePlot

2.2 Финансовые индикаторы

В Wolfram Mathematica финансовые индикаторы [3] представлены в виде функций, которые позволяют анализировать и интерпретировать финансовые данные. Встроенные функции обеспечивают широкий спектр индикаторов, которые могут быть использованы для анализа финансовых рынков, оценки акций и облигаций.

Полный список финансовых индикаторов в Wolfram Mathematica на рисунке 2.3:

```
FinancialIndicator[All]
Length[%]

{AbsolutePriceOscillator, AccumulationDistributionLine, AccumulativeSwingIndex, Aroon, AroonOscillator, AverageDirectionalMovementIndex,
AverageDirectionalMovementIndexRating, AverageTrueRange, ChaikinMoneyFlow, ChaikinOscillator, ChaikinVolatility, ChandeMomentumOscillator,
CloseLocationValue, CloseLocationValueVolume, CommodityChannelIndex, CommoditySelectionIndex, DemandIndex, DetrendedPriceOscillator,
DirectionalMovement, DynamicMomentumIndex, EaseOfMovement, FastStochastic, ForceIndex, ForecastOscillator, FullStochastic, Inertia,
IntradayMomentumIndex, KlingerOscillator, LinearRegressionSlope, MarketFacilitation, MassIndex, MESAPhase, MESASineWave, Momentum,
MoneyFlowIndex, MovingAverageConvergenceDivergence, MovingAverageConvergenceDivergenceHistogram, NegativeVolumeIndex, OnBalanceVolume,
PercentagePriceOscillator, PercentageVolumeOscillator, Performance, PolarizedFractalEfficiency, PositiveVolumeIndex, PriceVolumeTrend,
ProjectionOscillator, QStick, RandomWalkIndex, RangeIndicator, RateOfChange, RelativeMomentumIndex, RelativeStrengthIndex,
RelativeVolatilityIndex, RSquared, SlowStochastic, StandardDeviation, StandardError, StochasticMomentumIndex, SwingIndex, TimeSegmentedVolume,
TRIX, TrueRange, UltimateOscillator, VerticalHorizontalFilter, Volume, VolumeRateOfChange, WilliamsAccumulationDistribution,
WilliamsPercentR, AccelerationBands, BollingerBands, SimpleMovingAverage, HullMovingAverage, DoubleExponentialMovingAverage,
ExponentialMovingAverage, ParabolicStopAndReversal, VolatilitySystem, WeightedMovingAverage, LinearRegressionTrendlines,
LinearRegressionIndicator, MedianPrice, VariableMovingAverage, TriangularMovingAverage, TripleExponentialMovingAverage,
TimeSeriesForecast, KeltnerChannels, TypicalPrice, MovingAverageEnvelopes, PriceChannels, ProjectionBands, RaffRegressionChannel,
StandardDeviationChannels, StandardErrorBands, WeightedClose, WildersMovingAverage, Close, High, Low, Open, HighestHigh, LowestLow}
```

100

Рисунок 2.3 – Вывод списка всех финансовых индикаторов

Рассмотрим примеры некоторых из них:

Простая скользящая средняя (Simple Moving Average, SMA, рисунок 2.4) – это один из наиболее распространенных финансовых индикаторов, используемых для анализа временных рядов данных, включая цены активов, объемы торговли и другие финансовые показатели. Он помогает сгладить ценовые данные и выявить общую тенденцию или тренд. Задается формулой (2.1)

$$y_{n+i} = \frac{1}{n} \left(\sum_{k=i+1}^{n+i} x_k \right) \quad i = \overline{0, N-n}, \quad (2.1)$$

где y_i – это значение SMA в момент времени t_i , где x_k – это цена в момент времени t_i , N – размер исходного временного ряда, n – период. Начальные $n - 1$ значений временного ряда вычисляются с использованием частичных периодов.

```
TradingChart[{"EPAM", {{2023, 8, 1}, {2023, 8, 31}}}, {"SimpleMovingAverage"}]
```

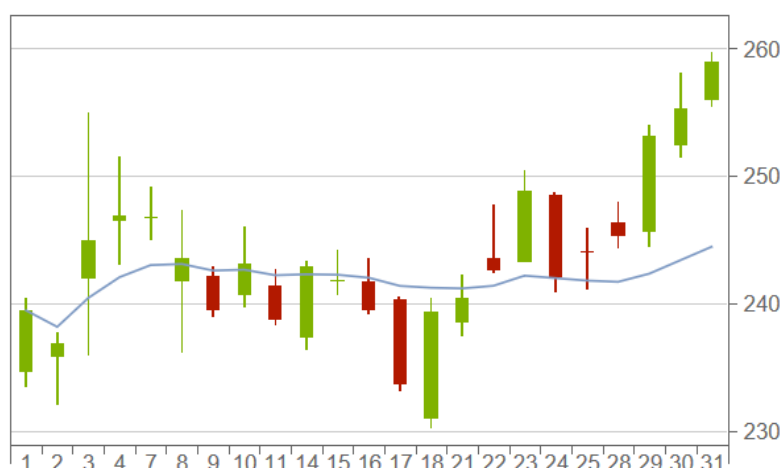


Рисунок 2.4 – Пример использования SimpleMovingAverage

ExponentialMovingAverage (экспоненциальная скользящая средняя, рисунок 2.5): Этот индикатор также сглаживает временной ряд, но с учетом весовых коэффициентов, которые уменьшаются экспоненциально с течением времени. Он часто применяется для быстрого реагирования на изменения тренда и задаётся следующей формулой (2.2)

$$y_{i+1} = y_i + \alpha(x_{i+1} - y_i), \quad i = \overline{1, n}, \quad (2.2)$$

где y_i – значение экспоненциального скользящего среднего в момент времени t_i , x_i – значение стоимости в момент времени t_i , α – коэффициент сглаживания обычно $\alpha \in [0, 1]$, но может быть любым выражением.

```
TradingChart[{"EPAM", {{2023, 8, 1}, {2023, 8, 31}}}, {"ExponentialMovingAverage"}]
```



Рисунок 2.5 – Пример использования индикатора ExponentialMovingAverage

BollingerBands (полосы Боллинджера, рисунок 2.6): Этот индикатор состоит из трех линий – центральной линии (скользящая средняя) и верхней и

нижней полос, которые находятся на некотором расстоянии от центральной линии и отражают стандартное отклонение данных. Он помогает определить периоды высокой и низкой волатильности и вычисляется с помощью 2-х формул (2.3) и (2.4).

$$u_i = m_i + \alpha \cdot \text{StandardDeviation}(n), \overline{0, N}, \quad (2.3)$$

$$l_i = m_i - \alpha \cdot \text{StandardDeviation}(n), \overline{0, N}, \quad (2.4)$$

где m_i – значения SMA(обычно с периодом 20) по формуле (2.1) в момент времени t_i , u_i – значение верхней границы в момент времени t_i , l_i – значение нижней границы в момент времени t_i , α – коэффициент определяющий ширину полос Боллинджера, обычно используют $\alpha = 2$, $\text{StandardDeviation}(n)$ – стандартное отклонение для периода n .

$$\text{StandardDeviation}(n) = \sqrt{\frac{1}{n} (\sum_{i=1}^N (x_i - \text{mean})^2)}, \quad (2.5)$$

где mean – среднее значение x_i , n – период.

`TradingChart[{"EPAM", {{2023, 8, 1}}, {2023, 8, 31}}], {"BollingerBands"}]`

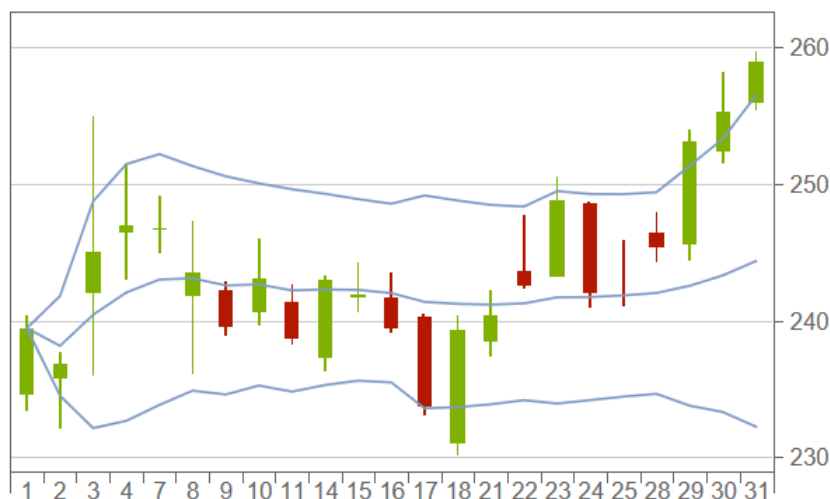


Рисунок 2.6 – Пример использования индикатора BollingerBands

MACD (стохастический осциллятор сигнальных линий, рисунок 2.7): Этот индикатор включает в себя две линии – MACD-линию и сигнальную линию. MACD-линия вычисляется как разность двух экспоненциальных скользящих средних с периодами 12 и 26, а сигнальная линия представляет собой экспоненциальную скользящую среднюю с 9 периодами MACD-линии.

Он помогает определить перекресты линий и сигнализирует о возможных изменениях тренда.

```
TradingChart[{"EPAM", {{2023, 8, 1}, {2023, 8, 31}}}, {"MACD"}]
```

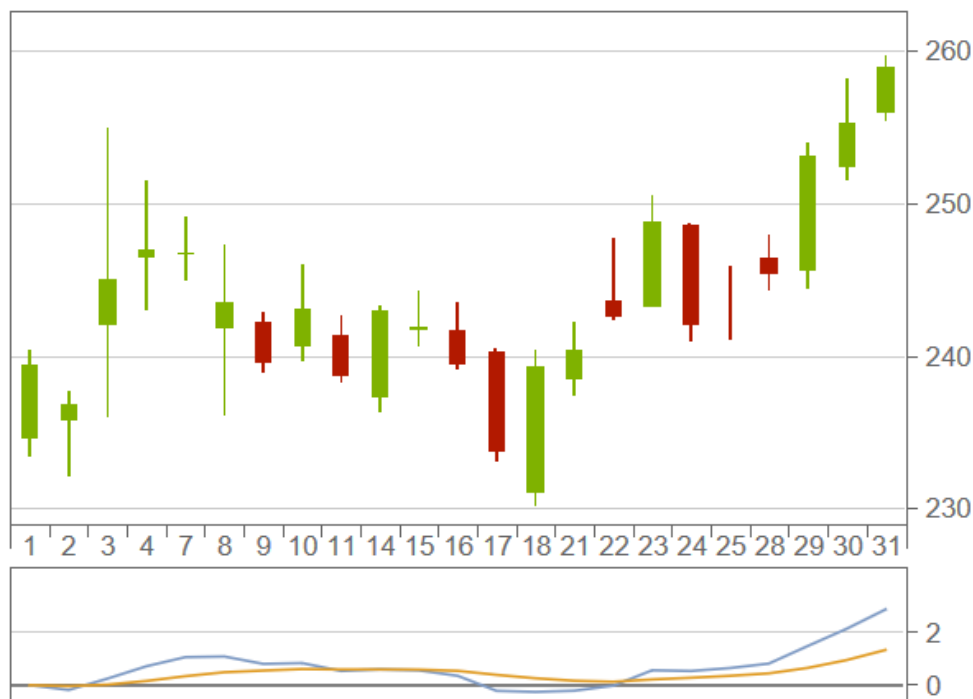


Рисунок 2.7 – Пример использования индикатора MACD

Набор графических объектов, предоставляемый системой Wolfram Mathematica, позволяет визуализировать данные различного объёма и структуры. Большое количество параметров графических объектов даёт пользователю гибкость в отображения данных, что несомненно является большим плюсом данной системы.

Глава 3

ОБЗОР БИБЛИОТЕК ДЛЯ ВИЗУАЛИЗАЦИИ И АНАЛИЗА ДАННЫХ НА ПЛАТФОРМЕ .NET

3.1 Общие сведения о платформе .NET

.NET — это мощная платформа разработки программного обеспечения, разработанная компанией Microsoft. Данная платформа предлагает широкий набор инструментов и библиотек для решения различных задач. Основными компонентами платформы являются:

- Common Language Runtime (CLR): CLR является средой выполнения платформы .NET, запускающая код написанный на таких языках как: C#, VB.NET, F#. Также данная среда обеспечивает управление памятью с помощью такого компонента как Garbage Collector (автоматический сборщик мусора) и повышает безопасность кода;
- .NET Core и .NET Framework. Эти компоненты предоставляют обширный набор реализованных классов, типов, методов для ускорения процесса разработки. Например, для любого сетевого приложения в пространстве имён System.Net найдётся множество полезных реализаций протоколов, слушателей (Listner) и т.д;
- Фреймворк ASP.NET. Данный фреймворк идеально подходит в случае, если вам необходимо реализовать динамическое веб-приложение, веб-сервис. Разработчики данного фреймворка включили в него множество полезных инструментов, позволяющие закрыть большое количество функционала коробочной настройкой. Например, логгирование, валидация приходящих запросов, обработка ошибок, фильтрация запросов настройка сетевых политик приложения и многое другое;
- Windows Presentation Foundation (WPF). WPF является улучшением его предшественника (WinForms). Упор в данном фреймворке сделан на более отзывчивый пользовательский интерфейс (GUI) и достигается это за счет поддержки сложной анимации, мультимедиа и доработанных элементов управления (контролов);
- Windows Forms (WinForms). WinForms является первым .NET фреймворком для создания десктоп приложений с упрощённым пользовательским интерфейсом нежели у его приемника WPF.

Для задачи визуализации данных из списка рассмотренных компонентов наиболее подходящим будет WPF. По стандарту фреймворк WPF предоставляет более 40 реализованных элементов управления(контролов), где каждый элемент управления имеет свою область применения. В силу того, что

перед нами стоит задача визуализации данных, то безусловно элемент управления Chart подходит нам наилучшим образом. Однако в ходе разработки нам может не хватить функционала, который предоставляется «из коробки» и в данной ситуации к нам на помощь идут пользовательские библиотеки, которые предоставляют некоторый специфический функционал заточенный под конкретную проблему. А удобный механизм Nuget-пакетов помогает нам очень просто и без лишней головной боли подключать их в проект.

3.2 Библиотека ScottPlot

ScottPlot — одна из множества пользовательских библиотек для визуализации данных, но имеющая ряд преимуществ среди себе подобных, а именно: отсутствие платной лицензии, полностью открытый исходный код и большой выбор реализованных графических объектов, не уступающих по качеству библиотекам с платной лицензией.

Рассмотрим несколько графических объектов, которые предоставляет нам библиотека ScottPlot [2]:

- **Линейный график (Line Plot):** Простейший линейный график, отображающий данные по заданному набору точек на плоскости. Идеально подходит для отображения временных рядов и математических функций в двумерном пространстве.
- **Свечной график (Candlestick Plot):** Самый популярный график, если речь заходит о визуализации ценового движения активов. По данному графику можно определять минимальную и максимальную стоимость акции, по цвету свечи – индикатор роста или падения акций.
- **Гистограмма (Histogram Plot):** График, который позволяет визуализировать распределение данных. Он разбивает значения на интервалы и отображает частоту или относительную частоту появления значений в каждом интервале.
- **Круговая диаграмма (Pie Chart):** График, который позволяет визуализировать долю каждого элемента в целом. Круговая диаграмма разделена на секторы, пропорциональные значениям, которые они представляют.
- **Баровая диаграмма (Bar Plot):** График, который отображает значения различных категорий в виде вертикальных или горизонтальных столбцов. Он может быть использован для сравнения данных между разными категориями.

Данные графические объекты наиболее частые для визуализации экономических данных.

Базовым элементом для отображения любого графика в данной библиотеке – Plot. Далее на этот элемент мы размещаем данные определённым образом, рисунок 3.1.

```
var plt = new Plot(600, 400);

// создаём образец данных X/Y
double[] xs = DataGen.Consecutive(51);
double[] ys = DataGen.Sin(51);

// создание линейного графика
plt.AddScatterLines(xs, ys, Color.Red, 3);

plt.SaveFig("scatter_lineplot.png");
```

Рисунок 3.1 – Пример кода для построения линейного графика

Основные настройки для объекта Plot:

- Height – высота графиком в пикселях (float)
- Width – ширина графика в пикселях (float)
- Palette – набор цветов используемых в графике (интерфейс IPalette)
- XAxis - ось на нижнем краю графика (класс Axis)
- XAxis2 – ось на верхнем крае графика (класс Axis)
- YAxis – левая вертикальная ось графика (класс Axis)
- YAxis2 – правая вертикальная ось графика (класс Axis)

Рассмотрим наиболее популярный график для визуализации данных OHLC, код на рисунок 3.2 и результат выполнения рисунок 3.3:

```
var plt = new ScottPlot.Plot(600, 400);

// генератор выборочных данных для моделирования ценовых данных с течением времени
OHLC[] prices = DataGen.RandomStockPrices(new Random(0), 60);

// Добавление на график финансовую диаграмму, используя массив объектов OHLC
plt.AddCandlesticks(prices);

plt.SaveFig("finance_quickstart.png");
```

Рисунок 3.2 – Пример кода для построения свечного графика



Рисунок 3.3 – Свечной график

3.3 Библиотека Skender.Stock.Indicators

Skender.Stock.Indicators - это пользовательская библиотека на языке C#, предназначенная для работы с техническими индикаторами на финансовых рынках. Библиотека предоставляет различные индикаторы, такие как скользящие средние, стохастический осциллятор, индекс относительной силы (RSI) и другие. Совместимость с фреймворком WinForms позволяет очень просто интегрировать этот инструмент в своё приложение. Рассмотрим несколько простых примеров получения значений индикаторов.

Рассмотрим получение таких же финансовых индикаторов, как и в второй главе (рисунки 3.4, 3.5, 3.6):

- Простая скользящая средняя (Simple Moving Average, SMA)
`var lookbackPeriods = 10;`

```
IEnumerable<SmaResult> smasValues = quotes.GetSma(lookbackPeriods);
```

Рисунок 3.4 – Получение значений индикатора SMA с периодом 10

- Экспоненциальная скользящая средняя (ExponentialMovingAverage)
`var lookbackPeriods = 10;`

```
IEnumerable<EmaResult> emasValues = quotes.GetEma(lookbackPeriods);
```

Рисунок 3.5 – Получение значений индикатора ЕМА с периодом 10

- Полосы Боллинджера (BollingerBands)

```
IEnumerable<BollingerBandsResult> bollingerBandsValues = quotes.GetBollingerBands();
```

Рисунок 3.6 – Получение значений индикатора BollingerBands

В данной библиотеке для получения значений какого-либо индикатора необходимо сформировать данные в виде коллекции объектов класса Quote имеющие следующие поля, рисунок 3.7

```
public class Quote : IQuote
{
    public DateTime Date { get; set; }
    public decimal Open { get; set; }
    public decimal High { get; set; }
    public decimal Low { get; set; }
    public decimal Close { get; set; }
    public decimal Volume { get; set; }
}
```

Рисунок 3.7 – Поля класса Quote

Далее у данной коллекции можно вызвать метод для вычислений значений определённого индикатора.

Фреймворк WPF и пользовательская библиотека ScottPlot показали отличный результат в решении задачи визуализации данных. Достигается это за счет широкого выбора графических объектов с богатым набором параметров для модификации, которые предоставляет библиотека ScottPlot. Но в сравнении с системой Wolfram Mathematica, платформа .NET уступает по трудозатратам. Если в системе Wolfram Mathematica уже встроено множество графических объектов и инструментов для работы с ними, то платформа .NET не имеет такого преимущества и существует необходимость поиска библиотек или разработка собственных, что многократно увеличивает трудозатраты.

Глава 4

ПРОГНОЗИРОВАНИЕ ФИНАНСОВЫХ ДАННЫХ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ

В условиях быстро растущих финансовых рынков, а параллельно с этим и увеличивающийся объём информации сильно затрудняет, а порой и вовсе делает невозможным, анализ и прогноз финансовых показателей человеческими силами, но это не беда, ведь мы живём во время, когда появились достаточные вычислительные мощности и множество разработанных методов машинного обучения, которые помогут нам строить прогнозы с заданной точностью, а также устанавливать взаимосвязи в большой выборке данных. На сегодняшний день для людей связанных с финансовыми данными и их изменением (аналитики, трейдеры, экономисты) такой инструмент незаменим. В данной главе мы рассмотрим ключевые подходы к прогнозированию финансовых временных рядов с использованием, как простых методов, так и усложнённых. В конце главы мы проведём сравнительный анализ рассмотренных нами методов.

4.1 Линейная регрессия

Линейная регрессия представляет собой фундаментальный метод, который лежит в основе многих более сложных алгоритмов машинного обучения. Несмотря на свою простоту, данный подход сохраняет свою актуальность за счет вычислительной эффективности. Также немало важно, что зачастую точность, которую обеспечивает данный метод, достаточна для заданных параметров. Основа данного метода – поиск линейной функции (4.1), описывающей зависимость целевой переменной (y) от признака (x)

$$y = kx + b, \quad (4.1)$$

где b – смещение модели, k – вес.

Если рассматривается метод множественной линейной регрессии, то формула (4.1) будет принимать вид (4.2)

$$y = f_{k,b}(x) = k_1x_1 + k_2x_2 + \dots + k_nx_n + b, \quad (4.2)$$

где b – смещение модели, k – вектор весов, x – вектор признаков.

Существует два основных способа обучения на основе алгоритма линейной регрессии:

- Прямое уравнение в аналитическом виде
- Снижение ошибки модели, на основе алгоритма градиентного спуска и его разновидностей

Прямое уравнение в аналитическом виде предлагает разрешить уравнение вида (4.3)

$$k = (X^T X)^{-1} X^T y, \quad (4.3)$$

где k – вектор весов, учитывающий смещение модели b

В данном уравнении, с увеличением числа признаков (увеличении размерности матрицы X) потребуются большие вычислительные мощности, поэтому данным метод менее предпочтителен, чем следующий.

Метод градиентного спуска, основная идея которого – это минимизация функции потерь, путем итеративного обновления весов в направлении, противоположном градиенту этой функции. Ниже разберём алгоритм. На этапе инициализации мы полагаем $k = (0, 0, \dots, 0)$ и ошибка модели $b = (0, 0, \dots, 0)$, на основе данных параметров строим прогноз по формуле (4.4)

$$\hat{y} = kx + b. \quad (4.4)$$

Рассчитываем градиенты для смещения и весов используя формулы (4.5) и (4.6)

$$\frac{\partial J(\theta)}{\partial b} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i), \quad (4.5)$$

$$\frac{\partial J(\theta)}{\partial k} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \cdot x_i, \quad (4.6)$$

где θ – вектор весов и смещений.

Используя полученные градиенты из (4.5), (4.6) обновим значения весов и ошибку модели по формулам (4.7) и (4.8)

$$k = k - \alpha \cdot \frac{\partial J(\theta)}{\partial k} \quad (4.7)$$

$$b = b - \alpha \cdot \frac{\partial J(\theta)}{\partial b} \quad (4.8)$$

Далее повторяем данные шаги пока изменение градиента не станет меньше заданного порога ε . С полученными k_{final}, b_{final} строим прогноз в виде (4.9)

$$y_{predict} = k_{final}x + b_{final}. \quad (4.9)$$

4.2 Дерево решений

Дерево решений представляет собой ещё один алгоритм машинного обучения, который использую не только для задач регрессии, но и для задач классификации [9]. По сути данный метод заключается в создании дерева, где корень дерева – это весь набор данных, внутренние узлы – это предикаты по результатам которых и происходит дальнейшее разбиение данных и листовые узлы – это конечные узлы представляющие собой результат, если мы говорим про задачи классификации, то результирующие значение – класс, если же задача предсказания, то результат – среднее значение целевой переменной для всех образов, который попали в данный узел.

Рассмотрим алгоритм построения данного дерева. На первом этапе нам необходимо выбрать критерий для разбиения данных. Для задач прогнозирования в качестве критерия используют среднеквадратическую ошибку (MSE) вычисляемую по следующей формуле (4.10)

$$\varepsilon = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (4.10)$$

Далее по выбранным нами признакам разбиваем наши данные и для каждого разбиения высчитываем критерий разбиения (в нашем случае это MSE) и на его основе делаем выбор конкретного разбиения. Для полученных новых групп данных мы повторяем действия, указанные выше, реализуем так называемую рекурсию до выполнения условий остановки. В итоге мы получаем дерево решений (модель) по которой и будем осуществлять прогноз новых данных.

4.3 Методы ансамблирования

Методы ансамблирования – представляют собой использование не одной модели для предсказания, а объединение нескольких для улучшения точности. Примеры метода ансамблирования – случайный лес и градиентный бустинг.

Случайный лес, представляет собой группу деревьев решений, которые обучаются параллельно, где итоговый результат – это среднее между всеми деревьями. Данный подход направлен на уменьшение разброса (бэггинг). Это особенно полезно в экономических приложениях, где данные могут быть шумными и неполными.

Градиентный бустинг – одна из передовых техник ансамблевого машинного обучения, принцип которого – это последовательное обучение более слабых моделей (за частую это деревья решений), где новая модель аппроксимирует остатки предыдущей модели.

Рассмотрим алгоритм градиентного бустинга. На этапе инициализации мы берём очень простую модель, например среднее значение целевой переменной (4.11)

$$\hat{y}^0 = \frac{1}{n} \sum_{i=1}^n y_i, \quad (4.11)$$

где y_i – истинные значения.

Далее необходимо вычислить остаток (ошибка прогнозирования) по формуле (4.12)

$$r_i = y_i - \hat{y}_i^{k-1}, \quad (4.12)$$

где r_i – остаток, $\hat{y}_i^{k-1}$ – предсказание модели с предыдущей итерации, y_i – истинные значения.

Далее на следующем шаге k по формуле (4.13) мы обучаем новую простую модель на остатках с прошлой итерации

$$\hat{y}^k = h_k(x) \quad (4.13)$$

Проедавав заранее заданное число итераций m , мы получаем итоговую модель (4.14)

$$\hat{y}^m = \hat{y}^0 + \sum_{i=1}^m \alpha h_i(x) \quad (4.14)$$

4.4 Сравнительный анализ методов

Получим экономические данные средствами Wolfram Mathematica и построим прогнозы данных используя методы, описанные в данной главе.

Далее, высчитав среднеквадратическую ошибку (MSE), проверим какой из методов дал наиболее точный прогноз.

```
df = Dataset[Table[<|"open" → row[[1]], "high" → row[[2]], "low" → row[[3]], "close" → row[[4]]|>, {row, pricesNumbers}]];
(*Линейная регрессия*)
linearModel = Predict[df[All, {"high", "low", "close"}] → df[All, "open"], Method → "LinearRegression"];
linearPred = linearModel[df[All, {"high", "low", "close"}]];
(*Регрессия случайного леса*)
rfModel = Predict[df[All, {"high", "low", "close"}] → df[All, "open"], Method → "RandomForest"];
rfPred = rfModel[df[All, {"high", "low", "close"}]];
(*Градиентный бустинг*)
gbModel = Predict[df[All, {"high", "low", "close"}] → df[All, "open"], Method → "GradientBoostedTrees"];
gbPred = gbModel[df[All, {"high", "low", "close"}]];
```

Рисунок 4.1 – Построение моделей рассмотренных методов

В данном коде (рисунок 4.1) мы на основе параметров “High”, “Low”, “Close” строим модели для прогноза параметра “Open”. Далее получим прогноз.

```
realData = Normal[df[All, "open"]]
(*Определение функции для вычисления среднеквадратичной ошибки*)
squaredMeanError[actual_, prediction_] := Total[(actual - prediction) ^ 2] / Length[actual];
linearError = squaredMeanError[realData, linearPred];
rfError = squaredMeanError[realData, rfPred];
gbError = squaredMeanError[realData, gbPred];
errors = {{linearError, rfError, gbError}};
results = Table[<|"Linear Regression" → error[[1]], "Random Forest" → error[[2]], "Gradient Boosting" → error[[3]]|>, {error, errors}];
results
```

Рисунок 4.2 – Подсчет ошибок для методов

Рассмотрим полученный результат, рисунок 4.3, 4.4.

```
{<|Linear Regression → 3.64982, Random Forest → 16.9741, Gradient Boosting → 1.26457|>}
```

Рисунок 4.3 – Результат среднеквадратических ошибок для рассмотренных моделей

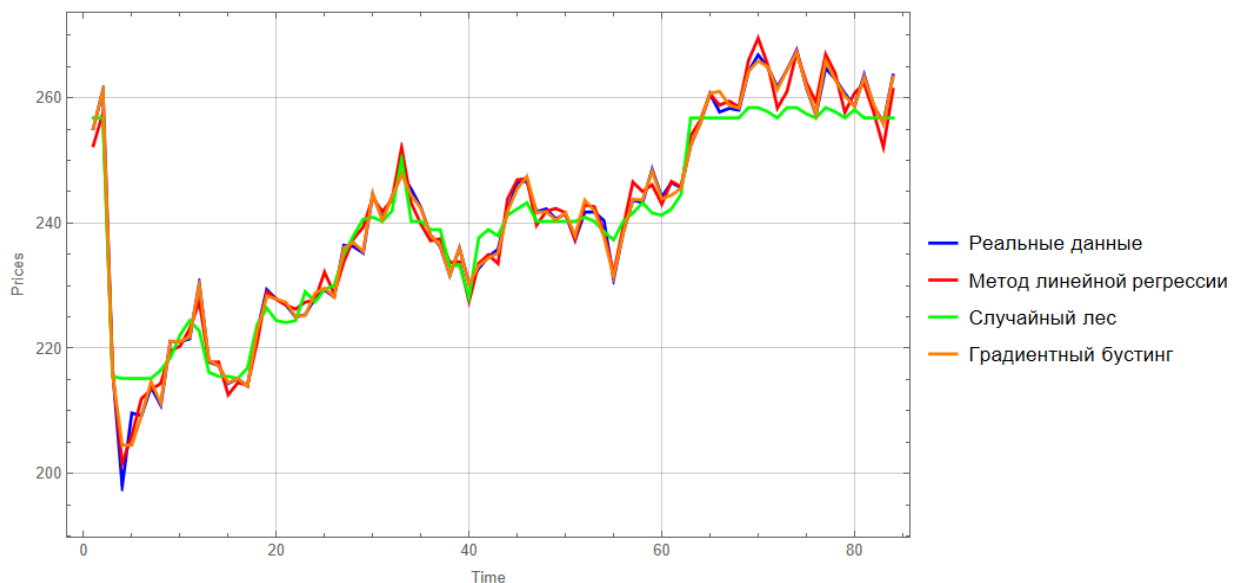


Рисунок 4.4 – Графическое сравнение полученных результатов в сравнении с реальными данными

Использование методов машинного обучения – это эффективный подход к решению задач прогнозирования, поэтому неудивительно, что система Wolfram Mathematica развивается в данном направлении. Описанные выше методы прогнозирования показали достойные результаты, но метод градиентного бустинга оказался самым эффективным. Решая реальные задачи прогнозирования финансовых данных, не всегда данный метод будет наилучшим, так как зависит от входных данных и порой будет достаточно более простых методов, как линейная регрессия, для построения прогнозов с заданной точностью.

Глава 5

СИСТЕМА DIRECTUM RX. ИНТЕГРАЦИЯ С АВТОМАТИЗИРОВАННОЙ ИНФОРМАЦИОННОЙ СИСТЕМОЙ «ВЕДОМСТВЕННЫЙ АРХИВ» (АИС ВА)

5.1 Системы электронного документооборота

Основной источник информации в организации – это документы (электронные и бумажные), создаваемые в рамках бизнес-процессов. Примерами бизнес-процессов могут быть: заявления на отпуск, подписание договора, составление отчетов, выдача справок в которых документ проходит через множество сотрудников организации, где у каждого своя задача в рамках бизнес-процесса. Для оптимизации работы с документами многие организации обращаются за помощью к системам электронного документооборота. Что же такое система электронного документооборота (СЭД)?

СЭД – программное обеспечение предоставляющие предоставляющее инструменты для работы с электронными документами на всех этапах их жизненного цикла: создание, редактирование, хранение, маршрутизация, поиск, классификация. В силу того, что документы и бизнес-процессы тесно связаны, то такие системы также предлагают широкий функционал и для работы с процессами: создание блок-схем, отслеживание текущего состояния, выбор исполнителей и это лишь самые распространённые функции.

На сегодняшний день многие организации озадачены вопросом перевода процессов и документов в цифровой формат, с чем безусловно и помогают СЭД, но независимо от сферы деятельности и размера бизнеса, есть ряд общих требований к таким системам:

- Защита информации;
- Обеспечение юридической силы документооборота;
- Возможность подписания документа электронной подписью;
- Интеграция с уже существующими системами, как внутренними, так и внешними;
- Масштабирование системы за счет прикладной разработки;
- Администрирование;

Переводя жизнедеятельность организации в цифровой формат за счет возможностей СЭД меняются следующие показатели:

- Сокращение трудозатрат сотрудников с работой с документами;

- Повышение исполнительской дисциплины. Каждый сотрудник видит полную картину текущего процесса, что позволяет ему грамотно распоряжаться временем выполнения поставленных задач;
- Повышение конкурентоспособности. Высокий уровень оптимизации и автоматизации внутренних процессов компании, позволяет предоставлять более качественные услуги своим клиентам, тем самым выделяя себя среди конкурентов;
- Сокращение расходов;
- Внедрение новых процессов в компании. Возможности доработки СЭД позволяет настроить практически любой новый процесс;

5.2 Система Directum RX

Своё видение СЭД предложила российская IT-компания Directum, занимающееся разработкой продуктов и сервисов для управления цифровыми процессами и документами. Исторически данная система относится к классу ECM (Enterprise Content Managment), что в переводе означает Управление корпоративным контентом, но на постсоветском пространстве, если речь заходит о ECM-системах, то зачастую говорят о СЭД.

Начало сегодняшней системе Directum RX было положено в 2002 году , но под другим именем – «Электронный документооборот». Данная система имела довольно узкий функционал, а именно «Электронный архив документов» и «Управление деловыми процессами», но не смотря на небольшой функционал системы, она успешно справлялась с большим количеством документов и пользователей. Следующим событием в истории развития системы было переименование системы «Электронный документооборот» на «Directum» и выбранный вектор развития системы (которому следуют и по сей день) – повышение эффективности процессов управления организацией и использование передовых технологий при работе с документами. Следующим «скачком» в возможностях системы «Directum» был релиз нового поколения – «Directum RX», обладающая следующими преимуществами над своим предком [5]:

- Обширный функционал в коробочном решении системы. Чего не было в предыдущих версиях;
- Добавлена интеллектуальная обработка документов, распознавание, классификация, авто регистрация документов, а также сравнение версий;

- Переработан интерфейс системы, что значительно повышает пользовательский опыт;
- В связи с развитием прикладной разработки, процесс адаптации стал значительно проще;
- Кроссплатформенность;
- Импортнезависимость. Используются инструменты с открытым исходным кодом.
- Развитие облачного варианта системы;

Рассмотрим архитектуру системы Directum RX, рисунок 5.1.

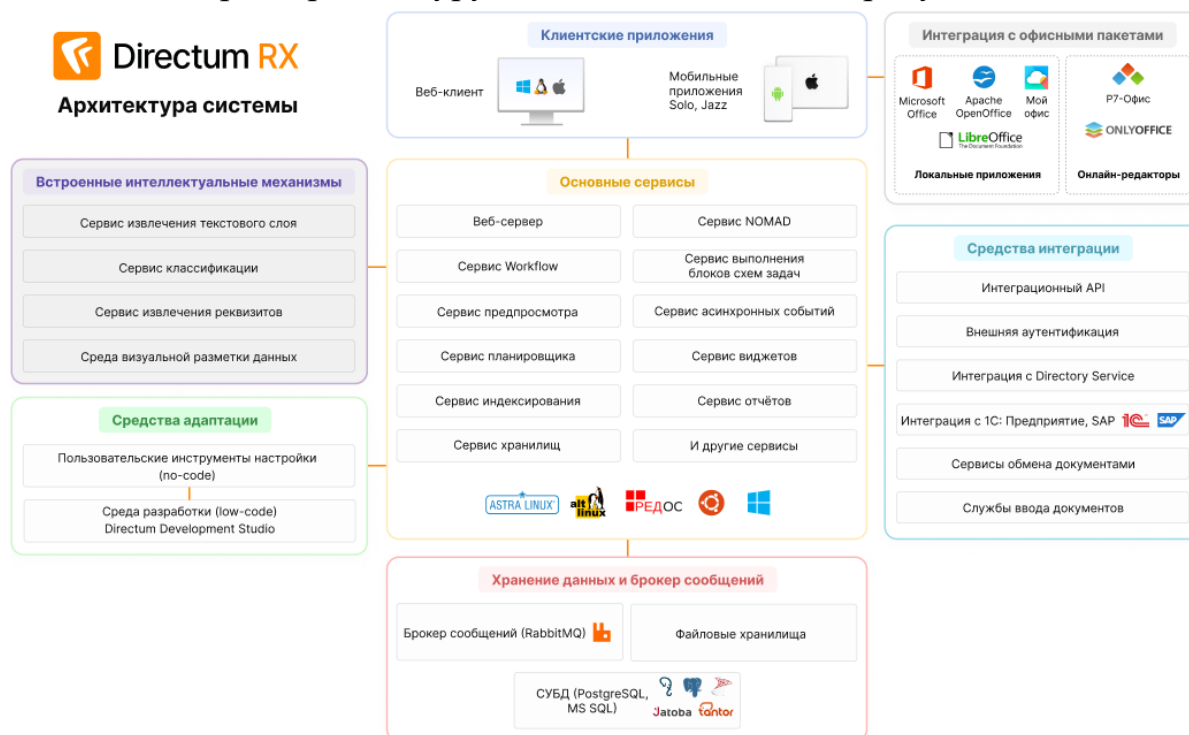


Рисунок 5.1 – Развернутая схема архитектуры Directum RX

Система имеет микросервисную архитектуру, где каждый сервис выполняет конкретную задачу и в случае загрузки конкретного сервиса, нет никаких проблем масштабировать её горизонтально (добавить вычислительную мощность) независимо от остальных сервисов, что позволяет строить мощное, масштабируемое решение для крупных организаций.

5.3 Автоматизированная информационная система «Ведомственный архив» (АИС ВА)

В условиях внедрения СЭД во многие организации вопрос о законном, надёжном и удобном хранении электронных документов актуален как

никогда. Именно для такой цели государственное предприятие «Национальный центр электронных услуг» (НЦЭУ) разработало автоматизированную информационную систему «Ведомственный архив» (АИС ВА) для архивной деятельности государственных органов и прочих организаций. Данная система представляет собой облачное решение, размещённое на собственной инфраструктуре предприятия. Такой подход избавляет организации от разворачивания собственной серверной инфраструктуры, а также от формирования команды технических специалистов, для её сопровождения. Доступ к системе осуществляется через привычный web-интерфейс, что делает взаимодействие с архивом простым и удобным.

Ключевым преимуществом данной системы является её соответствие законодательным требованиям РБ в области архивной деятельности [6]. Этот факт подкрепляется Указом № 188 Президента Республики Беларусь Александра Григорьевича Лукашенко от 31 мая 2022 года «О расширении использования государственными организациями информационно-коммуникационных технологий». Согласно 4 пункту данного Указа до 1 июня 2023 года республиканским органам государственного управления, подчиненным Правительству Республики Беларусь, облисполкомам и Минскому горисполкому необходимо обеспечить размещение и хранение электронных документов в автоматизированной информационной системе «Ведомственный архив» (АИС «Ведомственный архив»). Также система имеет Аттестат соответствия системы защиты информации информационной системы требованиям по защите информации (классы типовой информационной системы 3-фл, 3-юл в соответствии с СТБ 34.101.30-2017 «Информационные технологии. Методы и средства безопасности. Объекты информатизации. Классификация»), что говорит о надежности системы.

Широкий функционал системы охватывает весь жизненный цикл архивного электронного документа – с момента его приёма в систему и обработки до долговременного хранения и получения документа по запросу. Система работает с различными классами документов (описанными в справочнике ОКУД), форматами и объёмами. Каждый новый поступивший документ первостепенно проходит различные проверки: соответствие формату, подлинность электронной цифровой подписи (ЭЦП), права доступа, проверка целостности (в данной системе для этого используется стандарт симметричного шифрования под названием BelT) и другие.

Целостная структура архивных данных в АИС ВА – это результат добавления метаданных к каждому документу. С помощью этих метаданных строятся логические связи (в работе с документами речь идёт о группировке документов в дела). С выстроенной структурой разработчикам не составляло

труда, предоставить пользователю встроенную каталогизацию документов и поддержку фильтров по различным параметрам, что значительно ускоряет процесс поиска конкретного документа.

Огромным плюсом данной системы является возможность интеграции с внешними системами. Примером внешней системы может быть Directum RX, реализовав интеграцию с которой, организация полностью выстраивает процесс сквозной обработки документов – от их создания до архивного хранения. Такой функционал делает АИС ВА очень востребованной на рынке.

Обобщая всё вышесказанное о системе АИС ВА, если организация рассматривает переход от бумажных архивов, то лучше системы на территории Республики Беларусь, которая обеспечила бы такую же юридическую надёжность и законность хранения документов, на сегодняшний день нет.

5.4 Интеграция с АИС ВА

Основная цель этой интеграции – это внедрение функционала архивирования электронных документов АИС ВА в систему Directum RX. Итоговое решение позволит работать с электронными документами на всем их жизненном цикле – от создания до архивирования. Основными преимуществами внедрения решения в организации я бы выделил:

- Оптимизация работы персонала с электронными документами (время работы с документами значительно снижается);
- Переход от бумажного архива к электронному (снижение затрат на помещения и соответствующий персонал);
- Сокращение времени на получение доступа и поиск документов
- Удобство. Теперь сотруднику не требуется переключаться между разными web-приложениями, весь функционал теперь если в web-клиента системы Directum RX.

Для обмена сообщениями между данными системами используется протокол SOAP(Simple Object Access Protocol) позволяющий обмениваться структурированными данными в формате XML в распределённой среде, такой как интернет. Каждое SOAP-сообщение имеет общий скелет, а именно внешний тег <envelope> (конверт) и два вложенных тега <headers> и <body>. Такая стандартизация сообщений позволяет взаимодействовать абсолютно различным системам. Что касается транспортных протоколов, то здесь зачастую используется общепринятый протокол HTTP, но также возможно использование и таких протоколов как SMTP или JMS. Каждый транспортный

протокол упомянутый выше обеспечивает бесшовную интеграция разнородных систем. Также система АИС ВА поддерживает такой полезный механизм, как MTOM (Message Transmission Optimization Mechanism) основная задача которого – оптимизация как отправки, так и приёма бинарных данных. Внедрение данного механизма несёт ключевую роль в ситуациях, когда в систему загружаются большие объёмы бинарных данных(в реальность это может быть пакет с документами определённого дела, порой объём может достигать 200 мегабайт).

Перейдём к схеме взаимодействия с АИС ВА. Основной единицей обмена данными между интегрируемой ВСЭД (Внешняя система электронного документооборота) и АИС ВА – XML (Extensible Markup Language) пакет. Каждый XML-пакет должен быть составлен в соответствии с XSD (XML Schema Definition), в которой описан каждый тег: наименование, тип данных, частота вхождений тега, описание. Далее будем рассматривать каждый пакет его описание и назначение [6].

Первый пакет, который необходимо отправить для регистрации внешней системы (ВСЭД) в АИС ВА носит название XML-пакет «Регистрация». Ответом будет пакет, содержащий в себе уникальный идентификатор, по которому АИС ВА будет распознавать поступающие в неё пакеты. Структура пакета представлена на рисунках 5.2, 5.3 и 5.4.

1	2	3	4	5	6	
Envelope	–	–	–	–	–	1
–	Header	–	–	–	–	1
–	Body	–	–	–	–	1
–	–	RegistrationOrganization	–	–	–	1
–	–	–	Package	–	–	1
–	–	–	–	PackageBody	–	1
–	–	–	–	–	System	1
–	–	–	–	–	OrgInfo	1

Рисунок 5.2 – Уровни вложенности пакета «Регистрация»

sysid	1	String	Идентификатор ВСЭД. Для определения уникальности внешней системы (ВСЭД) при регистрации.	UUID / GUID – статистически уникальный 128-битный идентификатор, формируется на стороне системы ВСЭД.
sysname	1	String	Наименование системы отправителя (ВСЭД).	Произвольное наименование. При передаче разных пакетов рекомендуется формировать этот атрибут единообразно.
sysversion	1	String	Версия системы отправителя (ВСЭД).	Произвольное наименование. При передаче разных пакетов рекомендуется формировать этот атрибут единообразно.
url	1	String	Адрес внешней системы	Атрибут не обязательный к заполнению.

Рисунок 5.3 – Описание тега System

Имя допустимого атрибута	Кратность	Тип данных	Описание	Правила заполнения
nameOrg	1	String	Наименование организации.	Указывается наименование организации, до 200 символов.
unpr	1	int	Уникальный номер плательщика (УНП)	Атрибут должен содержать 9 цифр.
unpf	1	int	Уникальный номер плательщика фонда ФСЗН	Атрибут должен содержать 9 цифр.

Рисунок 5.4 – Описание тега OrgInfo

Пакет «Приемо-сдаточная опись», рисунок 5.5, необходим для уведомления системы АИС ВА о своих намерениях по передачи дел в архив. В составе данного пакета содержатся все метаданные по каждому делу описи, причем также есть информация по каждому документу. На данном этапе отправки система АИС ВА валидирует данные пакеты и отклоняет в случае ошибки, что значительно повышает производительность системы.

1	2	3	4	5	6	7	8	
Envelope		–	–	–	–		–	1
	Header							1
–	Body	–	–	–	–		–	1
–	–	SaveAcceptanceInventory	–	–	–		–	1
–	–	–	Package	–	–		–	1
–	–	–	–	Package Body	–		–	1
–	–	–	–	–	System		–	1
–	–	–	–	–	Inventory		–	
–	–	–	–	–	–	Document		1-n
–	–	–	–	–	–	Signature	–	1-n
–	–	–	–	–	–	Signature Time		1-n
–	–	–	–	–	–	EmployeeSubmittingDocuments	–	1
–	–	–	–	–	–		Position	1
–	–	–	–	–	–		Name	1
–	–	–	–	–	–	PrintInventory	–	0-1
–	–	–	–	–	–	–	File	1
–	–	–	–	–	–	–	FileName	1

Рисунок 5.5 – Уровни вложенности пакета «Приёмо-сдаточная опись»

XML-пакет «Статус приемо-сдаточной описи», рисунок 5.6, предназначен для получения данных о статусе приёмо-сдаточной описи, ранее переданной в АИС ВА. Возможные статусы в ответе FORMED – Сформирована, REJECTED – Отклонена, ACCEPTED – Принята. Разберём что значит каждый статус:

- FORMED – означает, что опись успешно получена АИС ВА и находится на проверке у архивиста АИС ВА;

- ACCEPTED – присваивается в случае пройденной проверки системы и успешно пройденном контроле архивистом АИС ВА;
- REJECTED – обозначает, что поступившая опись с метаданными дел и документов, не прошла контроль архивиста АИС ВА и удаляется из системы АИС ВА.

1	2	3	4	5	6	
Envelope	–	–	–	–	–	1
–	Body	–	–	–	–	1
–	–	GetInventoryStatusResponse	–	–	–	1
–	–	–	Package	–	–	1
–	–	–	–	ListInformation	–	1
–	–	–	–	–	Information	1
–	–	–	–	Inventory	–	0-1
–	–	–	–	–	Inventory ID	1
–	–	–	–	–	Status	1
–	–	–	–	–	Comment	0-1

Рисунок 5.6 – Уровни вложенности пакета «Статус приемо-сдаточной описи»

Ключевыми тегами в данном пакете является тег Document, в котором располагается XML-документ, рисунок 5.7, для описи в формате base64 и тег Signature, содержащий электронную цифровую подпись (ЭЦП) на приемо-сдаточной описи. Два этих тега несут основную информацию о планируемой загрузке описи. Рассмотрим метаданные, которые лежат в XML-документе описи.

1	2	3	4	5	6	7	8	
Inventory	–	–	–	–	–	–	–	1
–	StoreAttribute	–	–	–	–	–	–	1
–	Title	–	–	–	–	–	–	1
–	InDescription	–	–	–	–	–	–	1
–	Index	–	–	–	–	–	–	1
–	FundNumber	–	–	–	–	–	–	1
–	NumberKeeping	–	–	–	–	–	–	1
–	ListItems	–	–	–	–	–	–	1
–	–	ArchiveKeeping	–	–	–	–	–	1-n
–	–	–	Document	–	–	–	–	1
–	–	–	Hash	–	–	–	–	1

Рисунок 5.7 – Уровни вложенности в XML-документе для приемо-сдаточной описи

Здесь же ключевыми тегами являются тег Document содержащий XML-документ, описывающий дело, рисунок 5.8, включенное в опись, в формате base64 и тег Hash представляет собой строку фиксированной длины, которая подсчитывается по алгоритму BelT от содержимого тега Document. Корректность тега Hash напрямую влияет на целостность данных в системе. Рассмотрим схему XML-документа дела.

1	2	3	4	5
ArchiveKeeping		–	–	1
	IndexInInventory	–	–	1
–	Index	–	–	1
–	Title	–	–	1
–	BeginAkDocsTs	–	–	1
–	EndAkDocsTs	–	–	1
–	AmountED	–	–	1
–	VolumeED	–	–	1
–	Languages	–	–	1
–	–	Language	–	1-n
–	StoreAttribute	–	–	1
–	StorageTime	–	–	0-1
–	IndexSED	–	–	1
–	FundNumber	–	–	1
–	EPK	–	–	1
–	StorageComment	–	–	0-1
–	StorageArticle	–	–	0-1
–	OpenTs	–	–	1
–	ClosedTs	–	–	1
–	AmountFile	–	–	1
–	SubdivisionCode	–	–	1
–	RkkCount	–	–	1
–	Description	–	–	0-1
–	Note	–	–	0-1

Рисунок 5.8 – Уровни вложенности в XML-документе дела

Пакет «Внутренние описи», рисунок 5.9, необходим уже для передачи самих дел в архив, в нём содержатся все контрольные характеристики посылаемых в архив электронных документов дела (но на этом этапе пока без бинарных данных). Отличительной особенностью от других пакетов – это возможность отправки нескольких пакетов «Внутренние описи» для приемо-сдаточной описи. На данном этапе важно отследить, что кол-во заявленных документов в пакете «Приемо-сдаточная опись» будет совпадать с совокупностью файлов, отправленных во всех пакетах «Внутренние описи», если этого не соблюсти, то контрольная проверка не будет пройдена и ответственный за архивное хранения отклонит данную опись.

1	2	3	4	5	6	7	8	9	ость
Envelope	–	–	–	–	–	–	–	–	1
–	Header	–	–	–	–	–	–	–	1
–	Body	–	–	–	–	–	–	–	1
–	–	SaveInnerInventory	–	–	–	–	–	–	1
–	–	–	Package	–	–	–	–	–	1
–	–	–	–	Package Body	–	–	–	–	1
–	–	–	–	–	System	–	–	–	1
–	–	–	–	–	InnerInventories	–	–	–	1
–	–	–	–	–	–	InnerInventory	–	–	1-n
–	–	–	–	–	–	–	ArchiveKeepingID	–	1-n
–	–	–	–	–	–	–	Document	–	1
–	–	–	–	–	–	–	Signature	–	1
–	–	–	–	–	–	–	SignatureTime	–	1
–	–	–	–	–	–	–	OrgApprovePerson	–	1
–	–	–	–	–	–	–	–	Position	1
–	–	–	–	–	–	–	–	Name	1

Рисунок 5.9 – Уровни вложенности пакета «Внутренние описи»

Самую важную информацию тут в себе содержат теги Document (base64 строка от XML документа описывающий дело, собранный по предоставленной xsd-схеме, рисунок 5.10) и тег Signature, содержащий подпись от содержимого тега Document. Ниже описана xsd-схема документа дела.

1	2	3	4	5	
InnerInventory		–	–	–	1
–	CountDoc	–	–	–	1
–	CountFile	–	–	–	1
–	CountVolumeDoc	–	–	–	1
–	ListElectronicDocuments	–	–	–	1
–	–	ElectronicDocument	–	–	1-n
–	–	–	NumEDinArch	–	1
–	–	–	CreateTs	–	1
–	–	–	Title	–	1
–	–	–	CountFileED	–	1
–	–	–	VolumeFileDocument	–	1
–	–	–	PrevEdHashFiles	–	0-1
–	–	–	PrevEDHash	–	0-1
–	–	–	HashED	–	1
–	–	–	HashEDMetadata	–	1
–	–	–	NoteED	–	1

Рисунок 5.10 – Уровни вложенности XML-документа для внутренней описи

В XML-документе теги содержащие хэш высчитанный по алгоритму BelT все также служат для проверки целостности данных в системе. После

успешной отправки пакетов «Внутренние описи» система готова принимать уже непосредственно сами бинарные данные документов. Для этих целей предназначен XML-пакет «Документы», рисунок 5.11. Данный пакет также, как и пакет «Внутренние описи» может отсылаться несколько раз. На данном этапе также важно соблюсти кол-во передаваемых документов – оно должно быть равно количеству заявленных уже во внутренней описи. Также важно передавать корректные контрольные характеристики файлов и метаданных.

1	2	3	4	5	6	7	8	9	10	11	12	
Envelope												1
	Body											1
		SavePackageEO										1
			Package									1
				PackageBody								1
					System							1
					ElectronicObjectList							1
						ElectronicObjectDocument						1-n
							ArchiveKeepingId					1
							Document					1
							DocumentHash					1
							Files					1-n
								File				1
									FileName			1
									Volume			1
									Hash			1
									DocumentHashNumber			1
									CreateTs			1
									Source			0-1
									Encoding			0-1
									DigitalSignatures			0-n
										DigitalSignature		0-1
											Signature	0-1
											SignatureTs	0-1

Рисунок 5.11 – Уровни вложенности пакета «Документы»

В теге Document все также base64 строка, но уже от другой схемы, а именно XML-документ для электронного документа, рисунок 5.12. Ниже представлена его схема.

1	2	3	4	
ElectronicObject		—	—	1
—	AkIndexSED	—	—	1
—	NumberInKeeping	—	—	1
—	IdentificationCode	—	—	1
—	Title	—	—	1
—	Heading	—	—	1
—	OwnerName	—	—	1
—	ClassDoc	—	—	0-1
—	Languages	—	—	1
—		Language	—	1-n
—	HasRkk	—	—	1
—	NeedSanction	—	—	1
—	Keywords	—	—	1
—	ObjectType	—	—	0-1
—	Note	—	—	0-1
—	CreateTs	—	—	1
—	SourceApproveTs	—	—	1
—	FileCount	—	—	1
—	FileList	—	—	1
—	—	File	—	1-n
—	—	—	Name	1
—	—	—	Volume	1
—	—	—	HashCode	1
—	—	—	CreateTs	1
—	—	—	Encoding	0-1

Рисунок 5.12 – Уровни вложенности XML-документа для электронного документа

Заключительным пакетом в данной интеграции будет пакет «Признак окончания передачи данных по приемо-сдаточной описи». Его основная задача – это уведомление системы АИС ВА о том, что все электронные документы согласно переданной приемо-сдаточной описи переданы. Но в новейших версиях системы данный пакет не обязателен, за счет переноса данной логики в клиентское приложение.

На стороне ВСЭД необходимо будет реализовать модели под каждый отсылаемый пакет, а также модели под ответы от АИС ВА. Методы, которые конвертируют сущности системы Directum RX в необходимые сущности для пакетов. Что касается отправки объёмных документов, необходимо грамотно сформировать пакет и использовать необходимые заголовки в HttpClient для подключения вышеупомянутого метода оптимизации МТОМ, так как система АИС ВА в технической документации акцентирует внимание на том, что все документы свыше объёма в 90 мегабайт отправляются исключительно с помощью МТОМ.

Одной из трудностей в данной интеграции будет использование внешнего криптопровайдера Avest CSP – расширение криптографического ядра системы Microsoft Windows (разработанное в соответствии со спецификацией Microsoft Crypto API), которое позволяет использовать криптографические алгоритмы сертифицированные Государственным

Центром Безопасности Информации при Президенте Республики Беларусь (ГЦБИ) всем приложениям Microsoft Windows. Использование данного криптопровайдера необходимо для подписания электронных документов пользовательским сертификатом. В системе Directum RX для добавления такого функционала предусмотрен плагин AvestCryptoPlugin. Также помимо подписания, он позволяет генерировать хэш функцию по алгоритму BelT, необходимый для формирования контрольных характеристик отправляемых документов.

Рассмотрим схему взаимодействия системы Directum RX и АИС ВА, рисунок 5.13.

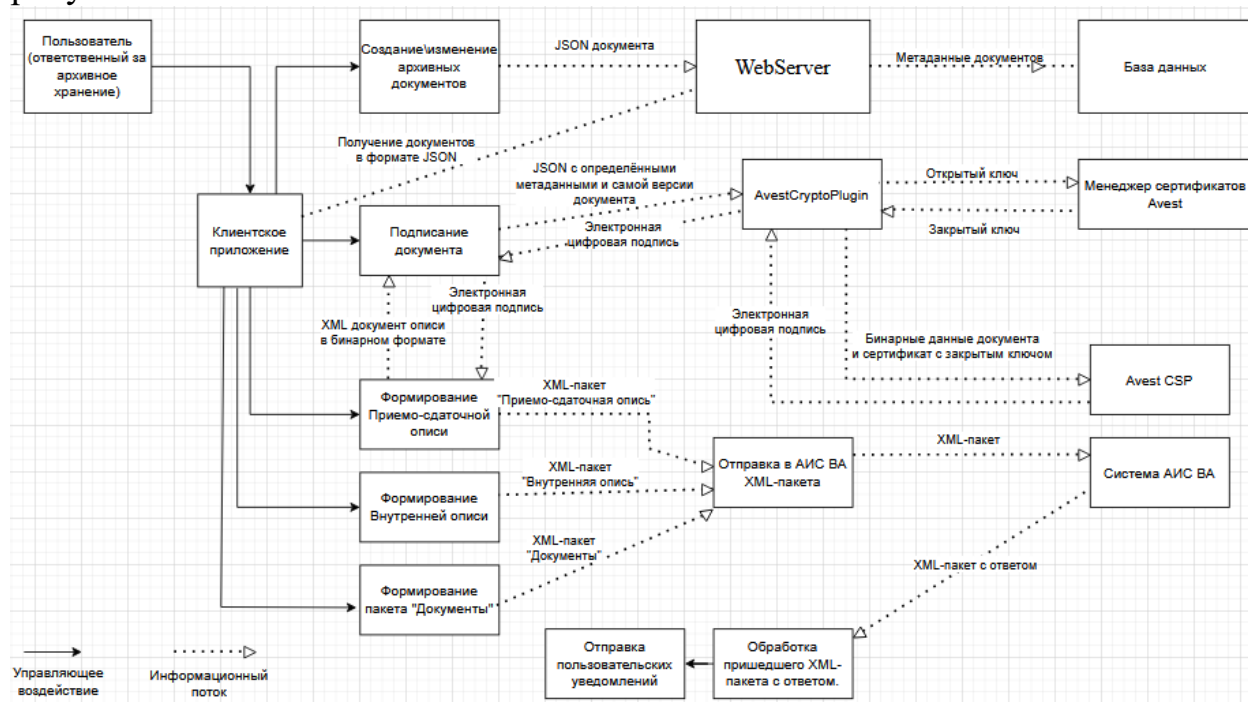


Рисунок 5.13 – Схема интеграции Directum RX с АИС «Ведомственный архив»

Разработанная интеграция системы Directum RX с АИС «Ведомственный архив» позволит организациям уйти от бумажного архивного хранения к цифровому, что ускорит процесс, как архивирования (ведь весь необходимый функционал будет находится в клиентском приложении Directum RX), так и извлечения необходимого документа из архива.

Глава 6

ПРИМЕРЫ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

В данной главе рассмотрим пример извлечения финансовых данных компании EPAM System Inc. с удалённого сервера используя средства и инструменты системы Wolfram Mathematica. Полученные данные, для наглядного понимания, визуализируем, используя графические объекты системы. После визуализации, составим прогноз ценовой динамики на 20 дней вперёд.

6.1 Извлечение и подготовка данных

В системе Wolfram Mathematica существует несколько вариантов извлечения данных:

- `FinancialData` – встроенная функция, позволяющая получать финансовые данные, такие как цены акций, валютные курсы
- `UrlRead` – встроенная функция, позволяющая взаимодействовать с различными внешними API по URL
- `Import` – встроенная функция для импорта данных в различных форматах (EXEL, JSON, XML, CSV)

Так как нас интересуют данные OHLCV компании EPAM в определённом временном промежутке, то наиболее удобным будет использование функции `FinancialData`.

Получим данные о стоимости акций компании EPAM Systems Inc. в период с 01.06.2023 по 01.10.2023, рисунки 6.1, 6.2.

```
financialData = FinancialData["EPAM", "OHLCV", {{2023, 06, 01}, {2023, 10, 01}}]
```

Рисунок 6.1 – Получение данных

```
convertedData = Map[Function[{item}, <|  
  "Date" → DateString[item[[1]], {"Day", "-", "Month", "-", "Year"}],  
  "Open" → QuantityMagnitude[item[[2, 1]]],  
  "High" → QuantityMagnitude[item[[2, 2]]],  
  "Low" → QuantityMagnitude[item[[2, 3]]],  
  "Close" → QuantityMagnitude[item[[2, 4]]],  
  "Volume" → QuantityMagnitude[item[[2, 5]] |>], financialData]  
Export["C:\\Users\\MI\\OneDrive\\Рабочий стол\\EPAMData.json", convertedData, "JSON"]
```

Рисунок 6.2 – Преобразование полученных данных в нужный формат и сохранение в формате JSON

6.2 Визуализация полученных данных

Для визуализации данных воспользуемся встроенными функциями системы Wolfram Mathematica, рисунки 6.4-6.6:

- TradingChart
- DateListPlot
- LineBreakChart

Так как при получении данных с удалённого сервера я использовал опцию "OHLCV", то для визуализации данных получим компоненты *Open, High, Low, Close, Volume*, рисунок 6.3.

```
openPrices = Map[{#[[1]], #[[2, 1]]} &, financialData]
highPrices = Map[{#[[1]], #[[2, 2]]} &, financialData]
lowPrices = Map[{#[[1]], #[[2, 3]]} &, financialData]
closePrices = Map[{#[[1]], #[[2, 4]]} &, financialData]
volume = Map[{#[[1]], #[[2, 5]]} &, financialData]
```

Рисунок 6.3 – Разбиение исходных данных на компоненты

```
DateListPlot[closePrices, PlotLabel → "EPAM Close Prices", PlotStyle → Blue, ImageSize → 600,
PlotLegends → {"Close price"}]
```

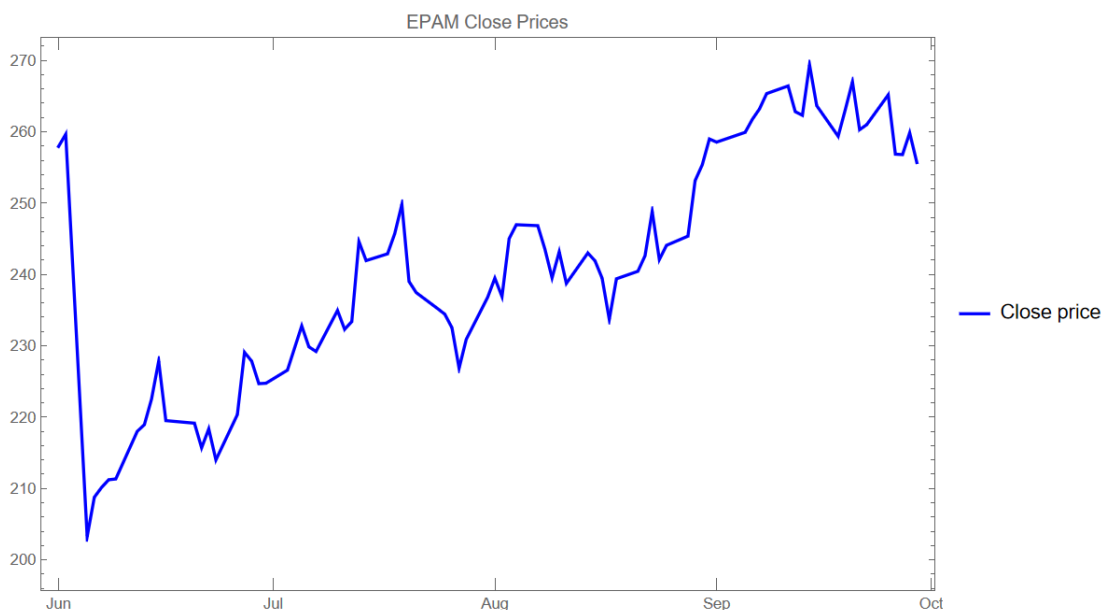


Рисунок 6.4 – Применение функции DateListPlot

```
TradingChart[financialData, PlotLabel → "EPAM Stock Prices", PlotTheme → "Detailed",  
ChartStyle → "Candlestick", ImageSize → 500]
```



Рисунок 6.5 – Применение функции TradingChart

```
LineBreakChart[closePrices, LabelStyle → "EPAM Close Prices", ImageSize → 500]
```

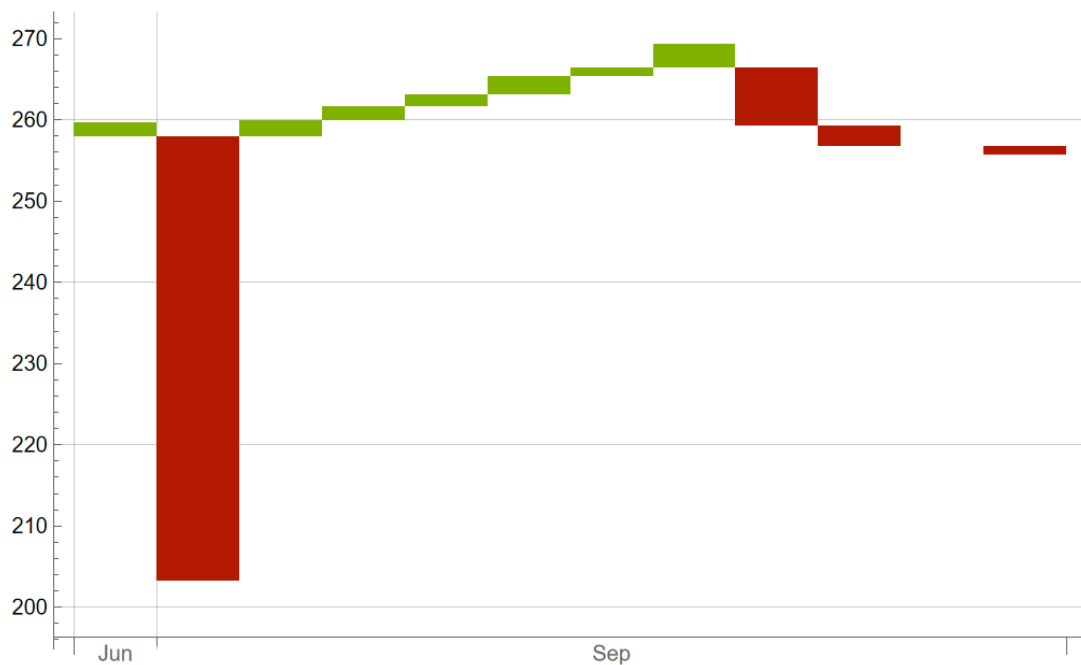


Рисунок 6.6 – Применение LineBreakChart

6.3 Расчёты прогнозов

Для демонстрации прогнозирования финансовых данных в системе Wolfram Mathematica используем следующие модели:

- Авторегрессионные модели (AR);
- Скользящее среднее (MA).

ARProcess (авторегрессионный процесс) представляет собой модель, в которой текущее значение временного ряда зависит от предыдущих значений этого же ряда. Авторегрессионная модель порядка p записывается следующим образом:

$$X_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t, \quad (6.1)$$

где X_t – текущее значение временного ряда, c – константа, $\varphi_1, \varphi_2, \dots, \varphi_p$ – коэффициенты авторегрессии, $X_{t-1}, X_{t-2}, \dots, X_{t-p}$ – предыдущие значения временного ряда, ε_t – случайная ошибка с нулевым средним и постоянной дисперсией.

MAProcess (скользящий средний процесс) представляет собой модель, в которой текущее значение временного ряда зависит от случайных ошибок (инноваций) предыдущих моментов времени. Скользящая средняя модель порядка q записывается следующим образом:

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \quad (6.2)$$

где X_t – текущее значение временного ряда, μ – среднее значение временного ряда, ε_t – случайная ошибка в текущий момент времени, $\theta_1, \theta_2, \dots, \theta_q$ – коэффициенты скользящей средней, $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ – случайные ошибки в предыдущих моментах времени.

Рассмотрим следующий код, рисунок 6.7:

```
source = FinancialData["EPAM", {{2023, 6, 1}, {2023, 10, 1}}, "Day"]];
DateListPlot[source, Joined → True, Filling → Bottom, ImageSize → 400]
FreeQ[source, Missing]
resampled = TimeSeriesResample[source, ResamplingMethod → {"Interpolation", InterpolationOrder → 0}]
RegularlySampledQ[resampled]
days = TimeSeriesResample[resampled, "Day"]
RegularlySampledQ[days]
ARmodel = EstimatedProcess[days, ARProcess[100]]
MAmodel = EstimatedProcess[days, MAProcess[100]]
forecastByAR = TimeSeriesForecast[ARmodel, days, {20}]
forecastByMA = forecast = TimeSeriesForecast[MAmodel, days, {20}];
realData = FinancialData["EPAM", {{2023, 10, 1}, {2023, 10, 20}}, "Day"]];
```

Рисунок 6.7 – Подготовка данных

В данном блоке кода мы прогнозируем финансовые данные на 20 дней вперёд с помощью двух моделей прогнозирования и визуализируем прогноз, рисунки 6.8, 6.9.

```
g1 = DateListPlot[{days , forecastByAR}, Joined → True, PlotStyle → {Gray, Magenta},
  ImageSize → 550, PlotLabel → "Прогноз на 20 дней с помощью ARmodel",
  PlotLegends → Placed[{"Реальные данные", "Прогноз"}, {Right, Bottom}]]
```

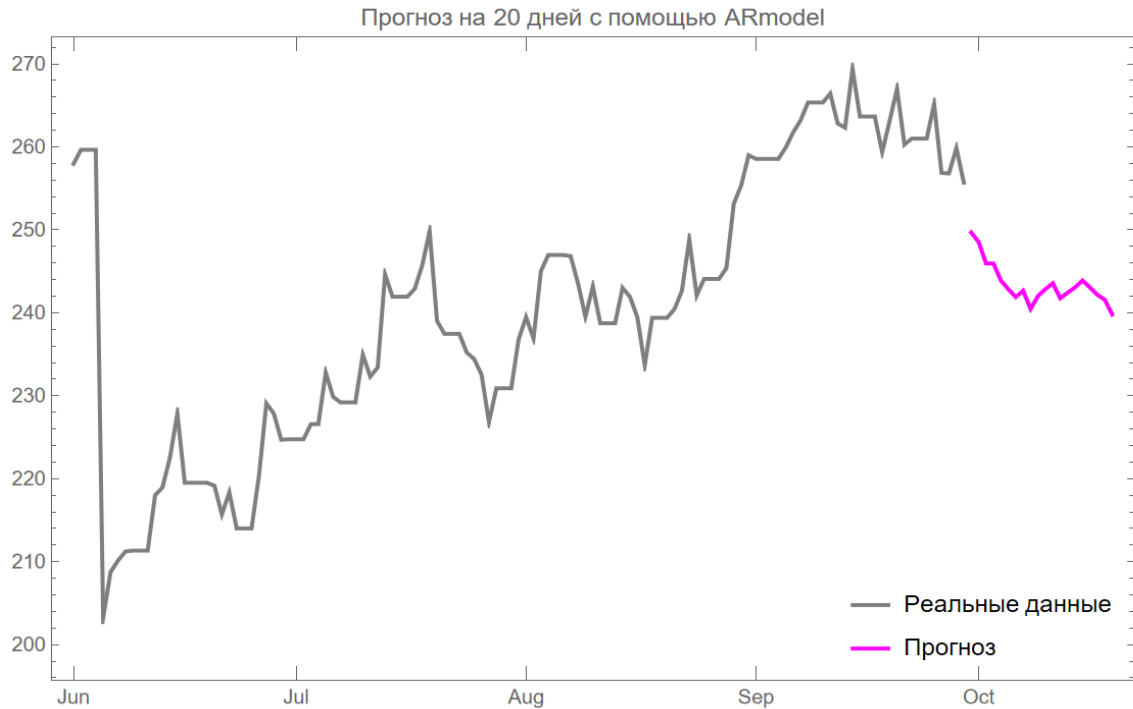


Рисунок 6.8 – Прогноз с применением AR модели

```
g1 = DateListPlot[{days , forecastByMA}, Joined → True, PlotStyle → {Gray, Magenta},
  ImageSize → 550, PlotLabel → "Прогноз на 20 дней с помощью MAmodel",
  PlotLegends → Placed[{"Реальные данные", "Прогноз"}, {Right, Bottom}]]
```

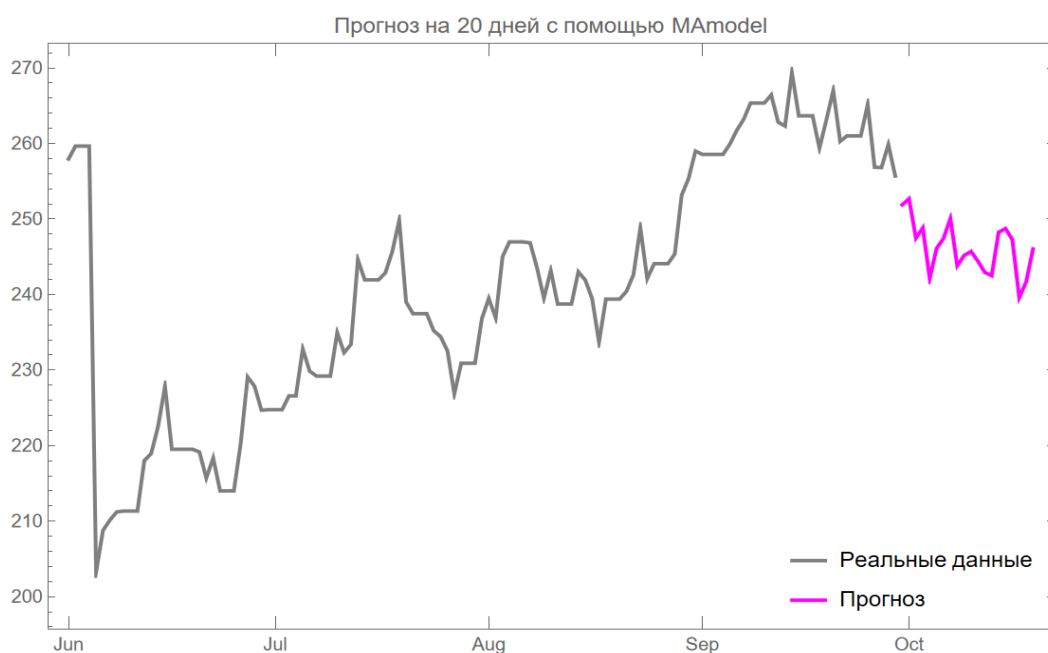


Рисунок 6.9 – Прогноз с применением МА модели

6.4 Сравнение прогнозов с реальными данными

Сопоставим прогноз и реальные данные, рисунки 6.10, 6.11.

```
g4 = DateListPlot[{days, realData}, Joined → True, PlotStyle → {Gray}, ImageSize → 550,  
PlotLabel → "Сравнение прогноза с реальными данными"]  
Show[g1, g4]
```

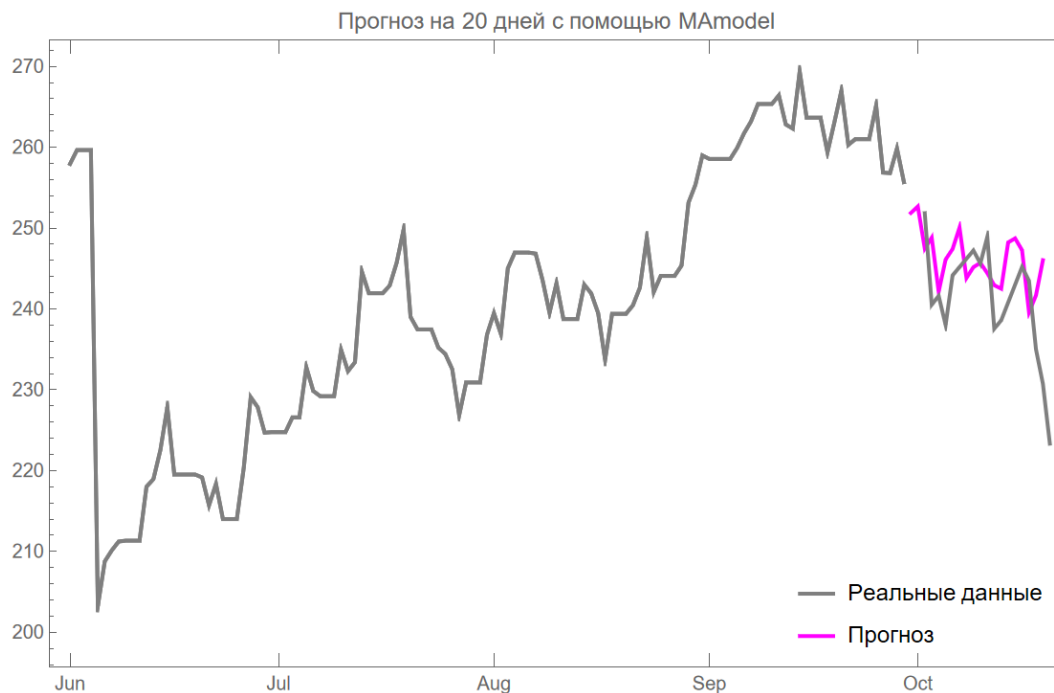


Рисунок 6.10 – Сравнение прогноза МА модели и реальных данных

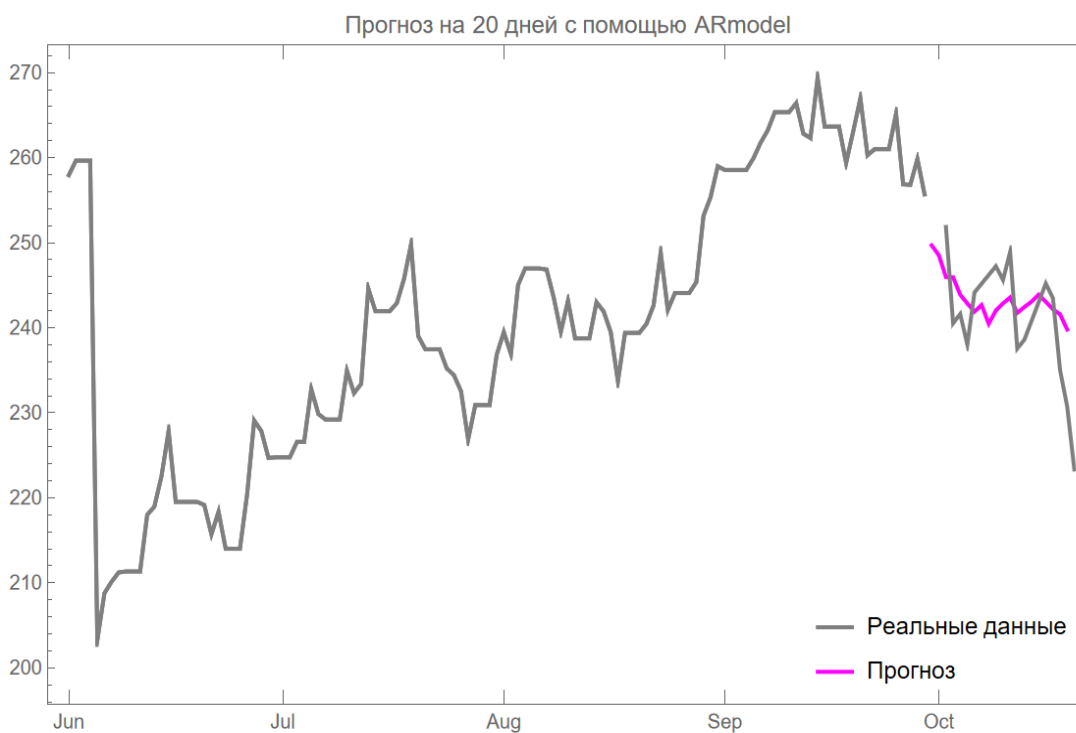


Рисунок 6.11 – Сравнение прогноза AR модели и реальных данных

6.5 Реализация на платформе .NET

Загрузим данные о стоимости акций компании ЕРАМ в период с 01.06.2023 по 01.10.2023, рисунок 6.12:

```
2 references
public IEnumerable<FinancialData> LoadFromFile(string path)
{
    if (!File.Exists(path))
    {
        throw new ArgumentException($"File with path={path} doesn't exists", nameof(path));
    }

    using (var reader = new StreamReader(path))
    {
        var text = reader.ReadToEnd();
        var data = JsonConvert.DeserializeObject<IEnumerable<FinancialData>>(text);
        return data;
    }
}
```

Рисунок 6.12 – Реализация метода для загрузки данных с файла формата JSON

Далее отобразим полученные данные в виде свечного графика (Candlestick Chart), рисунок 6.13 и его код на рисунке 6.14:



Рисунок 6.13 – Свечной график на платформе .NET

```

9 references
private void PrintCandlestickChart()
{
    var ohlc = _allData
        .Select(x => new OHLC(
            x.Open,
            x.High,
            x.Low,
            x.Close,
            x.Date,
            TimeSpan.FromDays(1)))
        .ToArray();

    mainPlot.Plot.Clear();
    mainPlot.Plot.AddCandlesticks(ohlc);
    mainPlot.Plot.XAxis.DateTimeFormat(true);
    mainPlot.Plot.XAxis.TickLabelFormat(x => DateTime.FromOADate(x).ToString("MMM dd, yyyy"));
    mainPlot.Refresh();
    UpdateVisibleBounds();
}

```

Рисунок 6.14 – Отображение загруженных данных

Для построения прогноза на платформе .NET будем использовать библиотеку Accord.Statistics, рисунок 6.15 и 6.16.

```

public IEnumerable<double> EstimateARModel(List<double> timeSeries, int order)
{
    var inputs = new double[timeSeries.Count - order][];
    var outputs = new double[timeSeries.Count - order];

    for (int i = order; i < timeSeries.Count; i++)
    {
        inputs[i - order] = new double[order];
        for (int j = 0; j < order; j++)
        {
            inputs[i - order][j] = timeSeries[i - j - 1];
        }
        outputs[i - order] = timeSeries[i];
    }

    // Оценка коэффициентов AR-модели с помощью OrdinaryLeastSquares
    var ols = new OrdinaryLeastSquares();
    var model = ols.Learn(inputs, outputs);

    var arCoefficients = model.Weights;

    return arCoefficients;
}

```

Рисунок 6.15 – Нахождение коэффициентов AR-модели

```

public IEnumerable<double> EstimateMAModel(List<double> timeSeries, int order)
{
    // подготовка входной матрицы
    var inputMatrix = new double[timeSeries.Count - order][];
    for (int i = 0; i < timeSeries.Count - order; i++)
    {
        inputMatrix[i] = new double[order];
        for (int j = 0; j < order; j++)
        {
            inputMatrix[i][j] = timeSeries[i + j + 1];
        }
    }

    var targetVector = new double[timeSeries.Count - order];
    for (int i = 0; i < timeSeries.Count - order; i++)
    {
        targetVector[i] = timeSeries[i];
    }

    // Оценка коэффициенты, используя обычный метод наименьших квадратов
    var ols = new OrdinaryLeastSquares();
    var model = ols.Learn(inputMatrix, targetVector);

    var coefficients = model.Weights;

    return coefficients;
}

```

Рисунок 6.16 – Нахождение коэффициентов МА-модели

С полученными коэффициентами мы строим прогноз на следующие 20 дней, рисунки 6.17, 6.18.

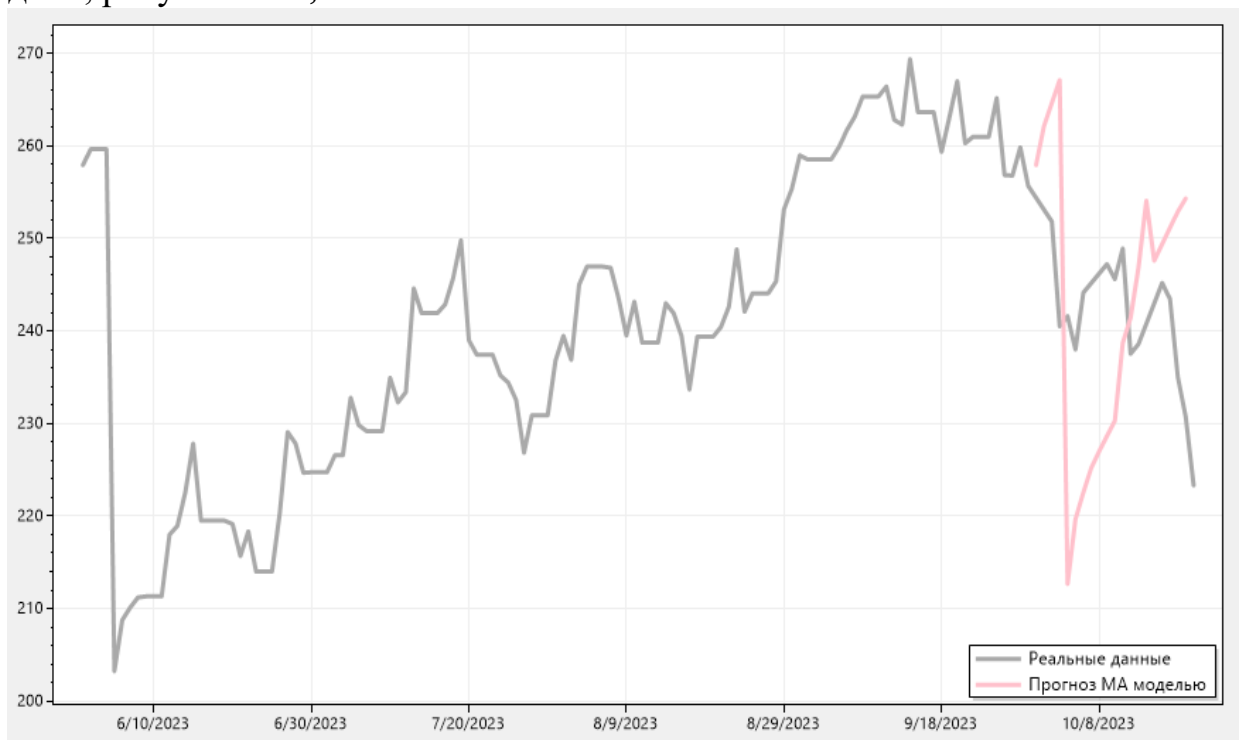


Рисунок 6.17 – Прогноз с применением МА модели

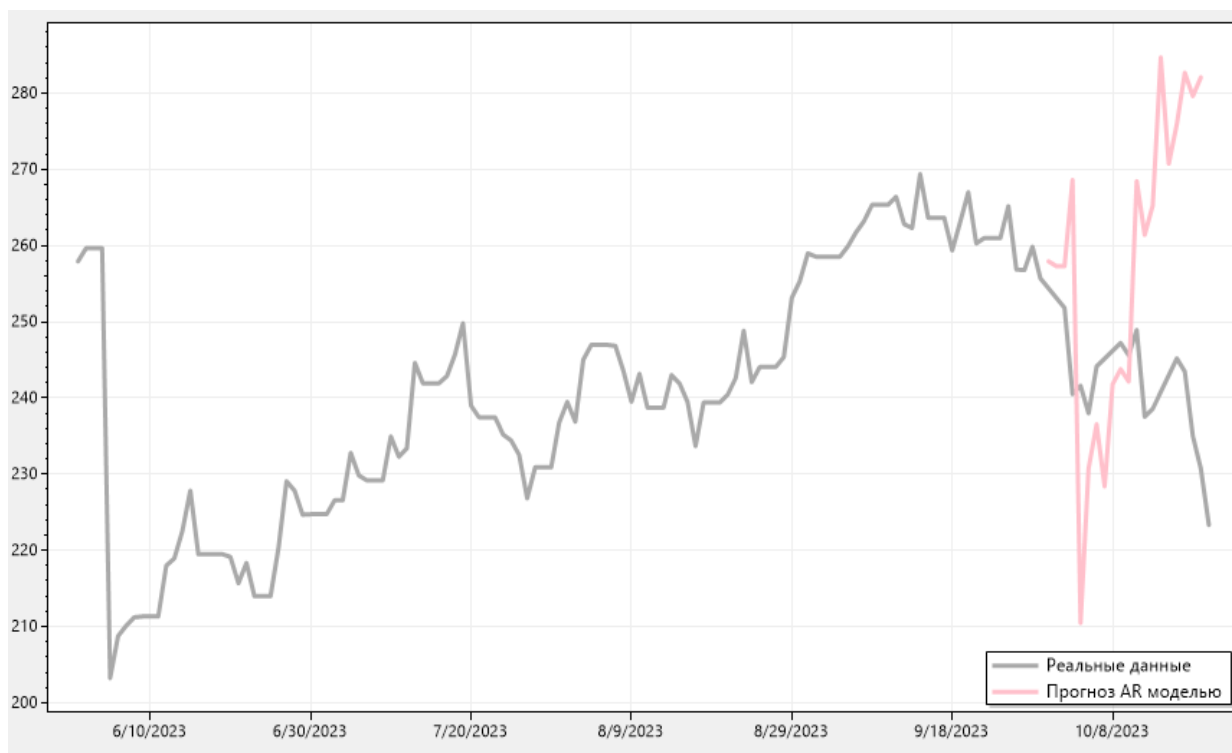


Рисунок 6.18 - Прогноз с применением AR модели

6.6 Сравнительный анализ полученных результатов

Для сопоставления результатов рассчитаем среднюю абсолютную ошибку по формуле (6.3)

$$\frac{1}{n} \cdot \sum_{i=1}^n |r_i - f_i| \quad (6.3)$$

где r_i – реальное значение, f_i – спрогнозированное значение.

В таблице 6.1 представлены результаты исследования.

Таблица 6.1 – Сравнения средней абсолютной ошибки в Wolfram Mathematica и .NET

Платформа	AR-модель	МА-модель
Wolfram Mathematica	4.3645680309494	4.88283383207125
.NET	25.2466339542445	16.4802614902246

Таким образом наглядно видно, что система Wolfram Mathematica значительно лучше справилась с задачей прогнозирования, чем реализация на платформе .NET. Что же касается количества кода, необходимого для решения данной задачи, то здесь как такового лидера не оказалось.

ЗАКЛЮЧЕНИЕ

Выполнение данной дипломной работы позволило погрузиться в процесс обработки, визуализации данных и методов её анализа в современных условиях. В первых главах дипломной работы рассмотрена необходимая теоретическая база об экономической информации, её классификации и способах анализа. Следующим этапом стало знакомство с системой Wolfram Mathematica, где детально рассмотрено многообразие графических объектов, которые предлагает система для решения задач визуализации. Как итог – система проявила себя с наилучшей стороны, за счёт большого количества специализированных графиков и возможности их детальной модификации, что в совокупности позволяет получать максимальную пользу от информации.

Альтернативным инструментом для решения задачи визуализации в дипломной работе была выбрана платформа .NET, в частности фреймворк WPF в паре с пользовательской библиотекой ScottPlot. Такой инструмент позволяет также успешно решать задачи визуализации, но с одним большим минусом – это необходимость поиска или длительной разработки библиотек с графическими объектами. Что же касается системы Wolfram Mathematica – никакого поиска и разработки, все уже имеется в «коробке» системы.

Одной из целей дипломной работы была оценка эффективности методов прогнозирования данных на разных платформах и их сравнительный анализ. В главах 4 и 6 детально рассмотрены методы прогнозирования данных их особенности и способы применения. Сравнительный анализ полученных результатов показал превосходство системы Wolfram Mathematica в точности построения прогнозов.

В рамках преддипломной практики затронута такая тема, как электронный документооборот, имеющую тесную связь с экономической информацией, ведь каждый документ организации – огромное количество полезных данных, которые можно использовать для анализа и оптимизации внутренних процессов. Реализованная интеграция с АИС «Ведомственный архив» – это востребованное решение для организаций различных размеров, в решении задачи перехода с бумажного архивного хранения документов на цифровую.

Таким образом, результаты работы подчеркивают важность современных инструментов анализа данных и их применения в управлении. Исследование системы Wolfram Mathematica показала, что она полностью покрывает весь цикл обработки экономической информации – извлечение, структурирование, обработка, визуализация, а также прогнозирование.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Баканов, М.И.* Теория экономического анализа: Учебник. / М.И. Баканов, А.Д. Шеремет - 4-е изд., доп. и перераб. – М. : Финансы и статистика, 2001. - 416 с.
2. Документация по библиотеке ScottPlot [Электронный ресурс] – Режим доступа: <https://scottplot.net>. – Дата доступа 02.05.2025.
3. Каталог «WOLFRAM Demonstrations Project». [Электронный ресурс]. – Режим доступа: <https://demonstrations.wolfram.com>. – Дата доступа : 27.11.2024.
4. Классификация видов экономического анализа [Электронный ресурс] – Режим доступа: <https://econ.wikireading.ru/69161>. – Дата доступа : 26.11.2024.
5. Официальный сайт компании Directum [Электронный ресурс]. – Режим доступа: <https://www.directum.ru>. – Дата доступа: 25.02.2025.
6. Официальный сайт Национального центра электронных услуг [Электронный ресурс]. – Режим доступа: <https://nces.by>. – Дата доступа: 05.03.2025.
7. *Таранчук, В.Б.* Введение в графику системы Mathematica : учеб. материалы для студентов фак. прикладной математики и информатики / В. Б. Таранчук. – Минск : БГУ, 2017. – 53 с.
8. *Таранчук, В.Б.* Одномерная графика системы Mathematica. Визуализация данных : учеб. материалы для студентов фак. прикладной математики и информатики / В. Б. Таранчук. – Минск : БГУ, 2021. – 48 с.
9. *Aileen Nielsen.* Practical Time Series Analysis. Released October 2019. O'Reilly Media, Inc.