МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ НАУК
Кафедра программного обеспечения

РЕКОМЕНДОВАНО К ЗАЩИТЕ В ГЭК
Заведующий кафедрой, к.т.н., доцент
_____М. С. Воробьева
02.07.
_____2021 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
магистерская диссертация

ПРИМЕНЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ НА ОСНОВЕ АВТОМАТИЧЕСКИХ ОТЧЕТОВ О ПРОДАЖАХ МАГАЗИНА НА ПЛОЩАДКЕ ETSY

02.04.03 Математическое обеспечение и администрирование информационных систем
Магистерская программа «Разработка технологий Интернета вещей и больших данных»

Выполнила работу
студентка 2 курса
очной формы обучения — Фокина Евгения Александровна

Научный руководитель
доцент кафедры программного
обеспечения, к. ф.-м. н. — Ступников Андрей Анатольевич

Рецензент
доцент кафедры программного
обеспечения, к. т. н. — Донкова Ирина Адольфовна

Тюмень
2021

ОГЛАВЛЕНИЕ

ОСНОВНЫЕ ТЕРМИНЫ

Листинг: страница магазина, имеющая уникальный номер и содержащая описание товара.

SKU: артикул товара.

Long-tail: «непопулярные» (то есть низкочастотные) и чётко сформулированные запросы, для которых характерна высокая конверсия.

Информационный критерий Акаике (AIC): критерий, применяющийся исключительно для выбора из нескольких статистических моделей. Разработан в 1971 как «an information criterion» («(некий) информационный критерий») Хироцугу Акаике и предложен им в статье 1974 года. Чем меньше значение, тем лучше модель.

ИАД: интеллектуальный анализ данных.

ВВЕДЕНИЕ

Продажа товаров ручной работы как основной или дополнительный источник дохода становится все более популярной. Как правило, изготовлением и продажей товаров ручной работы начинают заниматься в качестве хобби, и возможность получить доход является только дополнительным плюсом. Изделия ручной работы требуют большего времени для изготовления, в отличие от фабричных, более вдумчивого подхода к выбору материалов, и являются результатом в первую очередь творческого процесса. Мастера редко регистрируются как индивидуальные предприниматели или юридические лица, как правило, они регистрируются как самозанятые, то есть платят специальный налог на профессиональный доход без дополнительных отчислений подоходного налога или налога на прибыль.

Анализ литературы показал, что основные работы посвящены исследованию использованию информационных технологий для развития крупного и среднего бизнеса [1, 2, 3].

В России, в отличие от Запада, где товары ручной работы давно уже пользуются высоким спросом, ремесленническое сообщество только начинает складываться. Хендмейд-индустрия с одной стороны возрождает народные промыслы, а с другой - становится новым видом предпринимательства. Этому в немалой степени способствуют цифровые технологии и изменяющаяся психология потребителей, уставших от однообразия предлагаемых товаров и всё чаще обращающих внимание на товары ручной работы.

Несмотря на экономические трудности, проблемы взаимодействия с государственными органами, многие авторы развивают собственные бренды, создавая уникальные вещи и успешно реализуя результаты своего творчества. Чаще всего в России основным каналом продвижения являются соцсети, но в последнее время всё большую популярность завоёвывают торговые площадки, куда можно загрузить информацию об изделиях: «Ярмарка мастеров», Skafos, Ламбада Маркет, LOVE MADE [4].

Изменение внешней среды ведения бизнеса и усиление конкуренции заставляет многих мастеров искать дополнительные рынки сбыта своих товаров. Российский рынок товаров ручной работы оценивается в 30–60 млрд. рублей в год, в то время как мировой – в 20 млрд долларов [5]. Неудивительно, что российские мастера активно работают с иностранными покупателями, используя в своём бизнесе такие электронные торговые площадки как Etsy, Amazon Handmade, DaWanda, ArtFire, Zibbet, iCraft, RubyLane, Three Snails.

Наибольшей популярность среди российских хендмейкеров пользуется торговая площадка Etsy. Etsy – агрегатор, занимающийся электронной коммерцией и специализирующийся на винтажных товарах, товарах ручной работы и товарах для рукоделия. Сайт Etsy был запущен в 2005 году, и в настоящее время является ведущим онлайн-рынком товаров ручной работы по всему миру [6].

Несмотря на то, что на сайте Etsy присутствует возможность выбора языковых настроек, основным все-таки является английский. Большая часть пользователей – это жители Америки и Канады. Сервисы, ориентированные на анализ магазинов Etsy, так же в большинстве своем на английском языке.

Кроме того, человек, занимающийся творчеством, далеко не всегда знаком с особенностями ведения бизнеса, маркетингом, рекламой, что приводит к нерентабельности товаров, выставленных на продажу. Хендмейкеру, налаживающему бизнес на Etsy, зачастую проблематично разобраться в мануале компании: мешают сложности перевода, незнание законов маркетинга, рекламы и много другое. Нужен адаптированный для российского предпринимателя «помощник», способный упростить открытие и ведение магазина на Etsy.

В рунете можно найти достаточно инструкций и советов по оптимизации магазина на Etsy, однако они разрозненны и, как правило, касаются только какой-либо части процесса. Google-аналитика является прекрасным инструментом, однако ограничивается только анализом пользователей: география, демографические данные, конверсия и т.д.

Таким образом, у мастеров хендмейд индустрии существует потребность в комплексном инструменте для анализа экономических показателей магазина на

площадке Etsy. Наличие подобного инструмента существенно облегчит ведение бизнеса и поможет снизить финансовые расходы.

Целью выпускной квалификационной работы является разработка программного продукта для анализа и прогнозирования экономических показателей на основе автоматических отчетов о продажах магазина на площадке Etsy на основе применения методов интеллектуального анализа данных.

Для достижения этой цели были поставлены следующие задачи:

1. изучить существующие решения для анализа данных о продажах, ориентированные на работу с Etsy;
2. спроектировать процедуры обработки данных автоматических отчетов Etsy;
3. изучить методы обработки и визуализации статистических данных прогнозирования;
4. разработать приложение для анализа прогнозирования экономических показателей магазина на площадке Etsy.

Для успешной подготовки и защиты выпускной квалификационной работы обучающимся использовались средства и методы физической культуры и спорта с целью поддержания должного уровня физической подготовленности, обеспечивающую высокую умственную и физической работоспособность. В режим рабочего дня включались различные формы организации занятий физической культурой (физкультпаузы, физкультминутки, занятия избранным видом спорта) с целью профилактики утомления, появления хронических заболеваний и нормализации деятельности различных систем организма.

В рамках подготовки к защите выпускной квалификационной работы автором созданы и поддерживались безопасные условия жизнедеятельности, учитывающие возможность возникновении чрезвычайных ситуаций.

Для подготовки и защиты выпускной квалификационной работы использовались поиск, анализ информации, системный подход для решения поставленных задач; приемы критического анализа проблемных ситуаций, а также средства и методы саморазвития и самореализации; методики

межкультурного взаимодействия; умение расставлять приоритеты собственной деятельности при работе в общем проекте в соответствии с командной стратегией для достижения поставленной цели.

Формулирование выводов по итогам проведенной работы осуществлялись с учетом применения современных коммуникативных технологий (в том числе на иностранном языке) для представления результатов на академических, профессиональных, экспертных ИТ-мероприятиях.

ГЛАВА 1. ИНТЕЛЛЕКТУАЛЬНАЯ ОБРАБОТКА ДАННЫХ В ИНТЕРНЕТ-ПРОДАЖАХ

1.1. ОСНОВЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Математические методы обработки информации часто используются при проведении различных экономических исследований. Изучая бизнес с целью прогнозирования его будущего развития, необходимо разработать и изучить математические модели. Собранные статистические данные также необходимо обрабатывать и изучать. Исследования могут проводиться при помощи различных математических методов, но обработка больших объемов данных предполагает использование различных компьютерных технологий. Современные компьютерные технологии позволяют выходить на совершенно новый уровень обработки данных. «Интеллектуальная обработка информации» - это термин, который становится все более используемым. Интеллектуальная обработка данных предполагает использование не только новых инструментов, но и новых уровней математики, алгоритмов и программного обеспечения [7].

Понимание сущности анализа данных изменялось с развитием статистических методов обработки данных, и в настоящее время чаще всего говорят об интеллектуальном анализе данных (Data Mining) [8].

Data Mining (добыча данных, интеллектуальный анализ данных, глубинный анализ данных) – собирательное название, используемое для обозначения набора методов поиска в данных и интерпретации ранее неизвестной полезной информации, которая необходима для принятия решений в ряде областей человеческой деятельности. Термин введён Григорием Пятецким-Шапиро в 1989 году. Это процесс поддержки принятия решений, основанный на поиске в данных скрытых закономерностей (шаблонов информации). И в этом случае собранная информация автоматически преобразуется в информацию, которую можно идентифицировать как опыт.

Интеллектуальный анализ данных – это мультидисциплинарная область, возникшая и развивающаяся на базе таких наук как теория баз данных, статистика, искусственный интеллект, распознавание образов, машинное

обучение, алгоритмизация и другие. На рисунке 1 приведены основные дисциплины, на стыке которых появилась технология ИАД:

Рис. 1. Интеллектуальный анализ данных как мульти дисциплинарная область

В общем случае процесс ИАД состоит из трёх стадий:

- выявление закономерностей (свободный поиск);

- использование выявленных закономерностей для предсказания неизвестных значений (прогностическое моделирование);

- анализ исключений, предназначенный для выявления и толкования аномалий в найденных закономерностях.

Кроме того, часто между нахождением и использованием данных выделяют промежуточную стадию валидации, то есть проверки достоверности найденных закономерностей.

Выделяют 5 типов закономерностей, которые также называют задачами интеллектуального анализа данных [9]: ассоциация, последовательность, классификация, кластеризация и прогнозирование (или регрессия)

1. Ассоциация. При поиске решения проблемы в наборе данных обнаруживаются закономерности между связанными событиями. Специфика этой задачи заключается в том, что поиск закономерностей производится между разными событиями, происходящими в одно и то же время. Алгоритм Apriori является наиболее известным алгоритмом решения задачи поиска ассоциативных правил.

2. Последовательность. Последовательность похожа на ассоциацию, но ее главной целью является поиск закономерностей, происходящих через определенные интервалы времени. Ассоциацию можно рассматривать, как частный случай последовательности с временным лагом, равным нулю. Правило последовательности заключается в том, что после события *А* через определенное время наступит событие *Б*.

3. Классификация. Задача классификации состоит в том, чтобы разделить множество объектов на отдельные группы по схожим характеристикам. Эти группы называют классами. Это выполняется при помощи анализа значений признаков (или атрибутов) объектов. Для классификации используется множество различных моделей, такие как нейронные сети, деревья решений, алгоритмы покрытия, опорные вектора и другие.

4. Кластеризация. Эта задача является более сложным вариантом классификации. Особенность кластеризации заключается в том, что классы объектов не определены, разбиение на группы является результатом процесса.

5. Прогнозирование (регрессия). Этим методом изучаются тенденции развития различных процессов на основе анализа их изменений во времени.

При разработке приложения были решены задачи поиска ассоциативных правил для анализа потребительской корзины и определения совместно покупаемых товаров и прогнозирования временных рядов для предсказания уровня дохода на определенный период и пиков активности покупателей.

1.2. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ В ПРОДАЖАХ

Data Mining – это удобный инструмент для исследования продаж. С его помощью можно решить множество разных задач для компаний, чья деятельность связана с розничной торговлей. К таким задачам относятся:

- Прогнозирование на основе временных рядов. Используя архивные данные продаж компании, применение данного анализа позволяет отследить тенденции и сделать прогнозы на определенные периоды.

- Анализ рыночной корзины. Исследуя данные о структуре и сумме чека, данный анализ позволяет отследить, какие товары покупаются друг с другом, какие товары пользуются наибольшим спросом, а какие наименьшим. Основываясь на этих данных можно оптимизировать раскладку товаров, разработать акционные предложения и так далее.

- Формирование ассортимента и анализ продаж товаров. Исследуя данные о продажах товаров можно построить подробные профили различных продуктов и категорий продуктов. Найдя типичные закономерности в продажах товаров, можно более подробно изучить целевую аудиторию, что, в свою очередь, позволить сосредоточиться на тщательном продвижении определенных групп товаров.

- Анализ профилей поставщиков и покупателей. Это позволяет эффективно распределять затраты на маркетинг, проводить анализ доходности, увеличивать лояльность клиентов и поставщиков, исследовать различные способы привлечения новых клиентов и удержания уже существующих.

Разрабатываемое приложение решает некоторые из выше перечисленных задач, такие как прогнозирование и анализ корзины. Кроме того, частично решает задачу анализа продаж товаров и формирования ассортимента. Имеющиеся данные позволяют также работать над задачами анализа профиля клиента и целевой аудитории, однако данная функция в приложении реализована не будет.

1.3. АНАЛИЗ СУЩЕСТВУЮЩИХ РЕШЕНИЙ

Среди приложений для маркетингового анализа существует несколько решений, ориентированных на пользователей Etsy. В таблице 1 представлены результаты сравнения этих приложений по некоторым параметрам.

Таблица 1

Сравнение функций приложений, интегрированных с сайтом Etsy

Функции приложения	**eRank**	**Marmalead**	**Vela**	**Putler**	**Craftybase**
SEO аналитика	+	+			
Анализ продаж				+	
Контроль стока продукции					+
Контроль стока материалов					+
Обработка изображений			+		
Редактирование листингов непосредственно из программы			+		
Массовое редактирование			+		
Ведение бухгалтерии					+
Стоимость	6-10	19$		30-250	9-25
Пробный период				14 дней	14 дней
Бесплатная версия	+		+		

Сервис eRank

Сервис eRank [10] предлагает SEO анализ магазина. Ежемесячно формирует отчет о самых популярных поисковых запросах в Etsy, Amazon, eBay,

Pinterest и Google Shopping. Предлагает три уровня тарифных планов в зависимости от количества предоставляемых услуг: бесплатно, 6 долларов и 10 долларов в месяц. На рисунке 2 показан пример пользовательского интерфейса сервиса eRank.

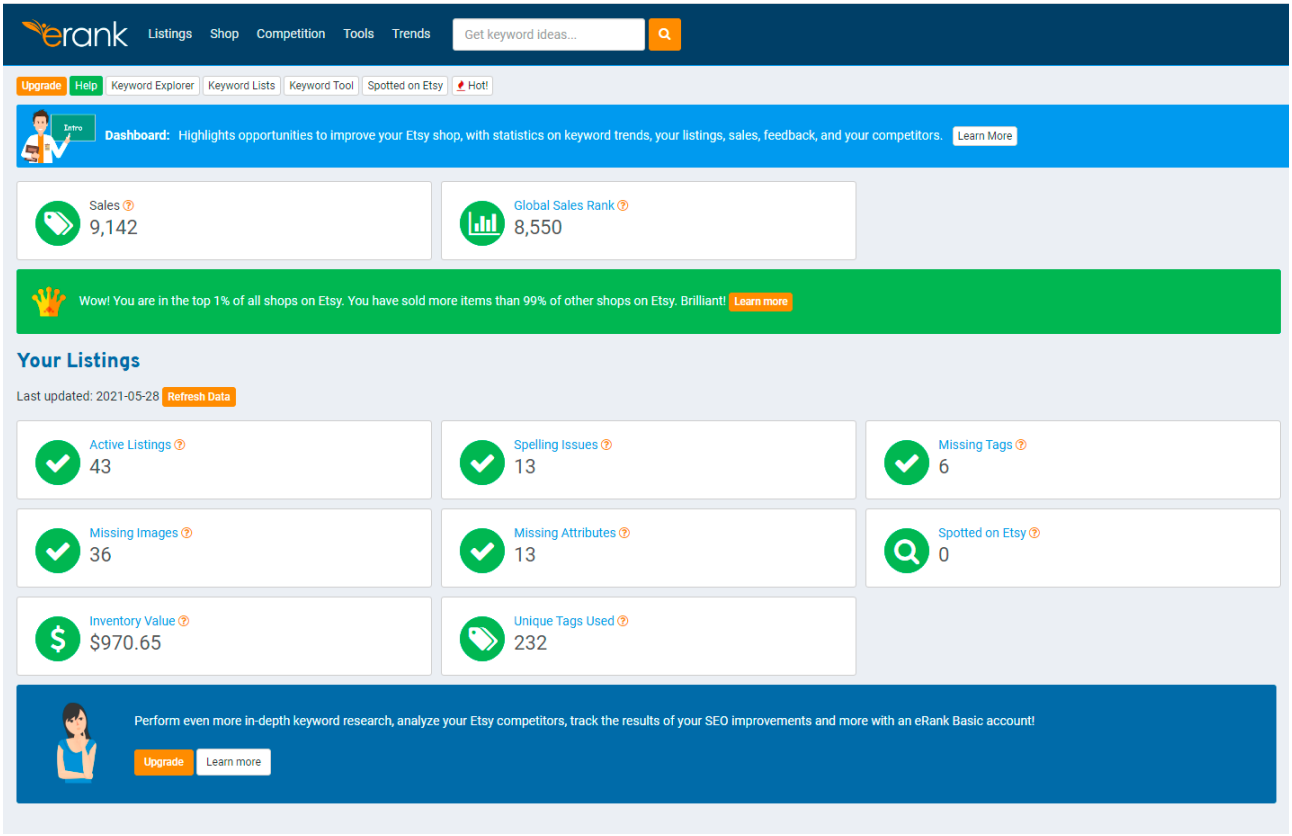

Рис. 2. Пример пользовательского интерфейса сервиса ERANK

Сервис Vela

Сервис Vela [11] позволяет централизованно анализировать и редактировать все листинги в нескольких магазинах Etsy. Используя это приложение, можно изменить цену товаров на определенную сумму или процент, опубликовать уведомления во описаниях товаров, создавать и редактировать профили вариаций. Vela ориентирована в первую очередь именно на массовое редактирование листингов и работу с фотографиями. Сервис предоставляется бесплатно. На рисунке 3 показан пример пользовательского интерфейса сервиса Vela.

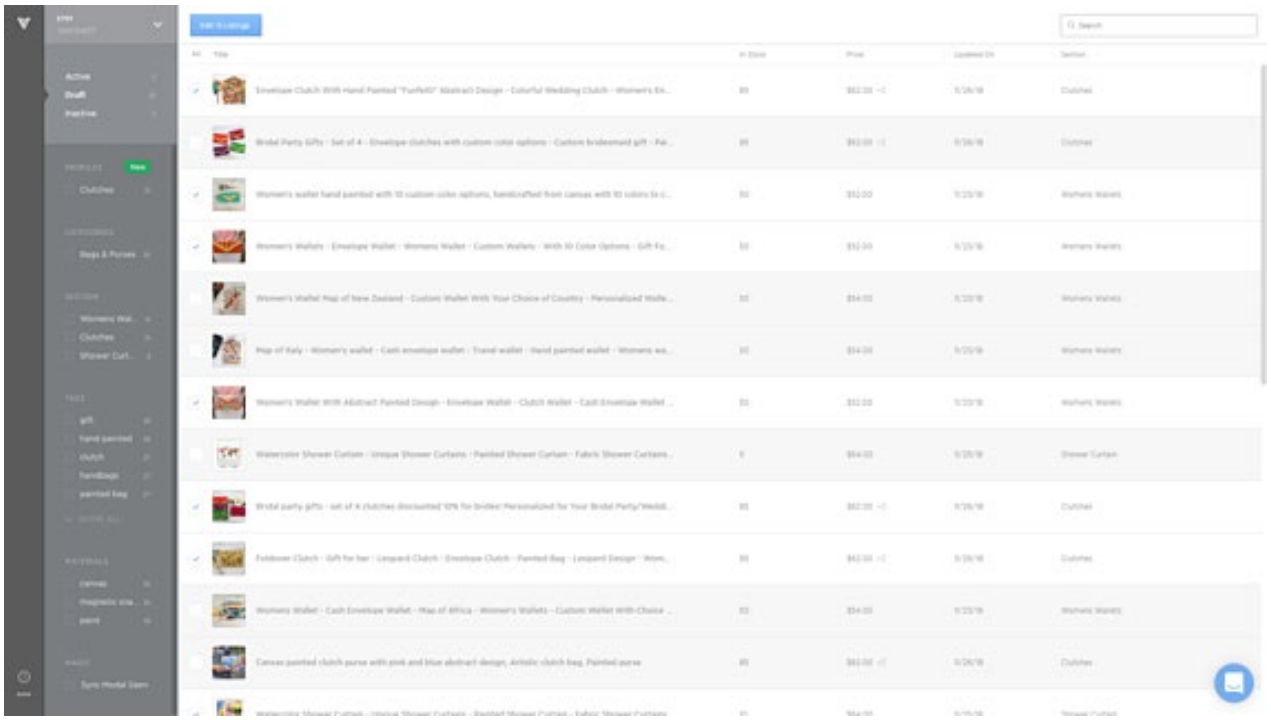

Рис. 3. Пример пользовательского интерфейса сервиса Vela

Сервис Craftybase

Сервис Craftybase [12] предназначен для работы с инвентарем и ведения бухгалтерии, специально разработан для отслеживания производства и продажи изделий ручной работы. Предлагает отслеживание затрат на материалы и продукцию, продаж и рентабельности. Позволяет контролировать уровень запасов и создавать «рецепты изделия», можно также установить оповещения о низком запасе материалов. Имеет еще несколько интересных функций, но ориентирован, в первую очередь, на резидентов США. Стоимость от 9 до 25 долларов в месяц в зависимости от тарифного плана. Бесплатный период – 14 дней. На рисунке 4 показан пример пользовательского интерфейса сервиса Craftybase.

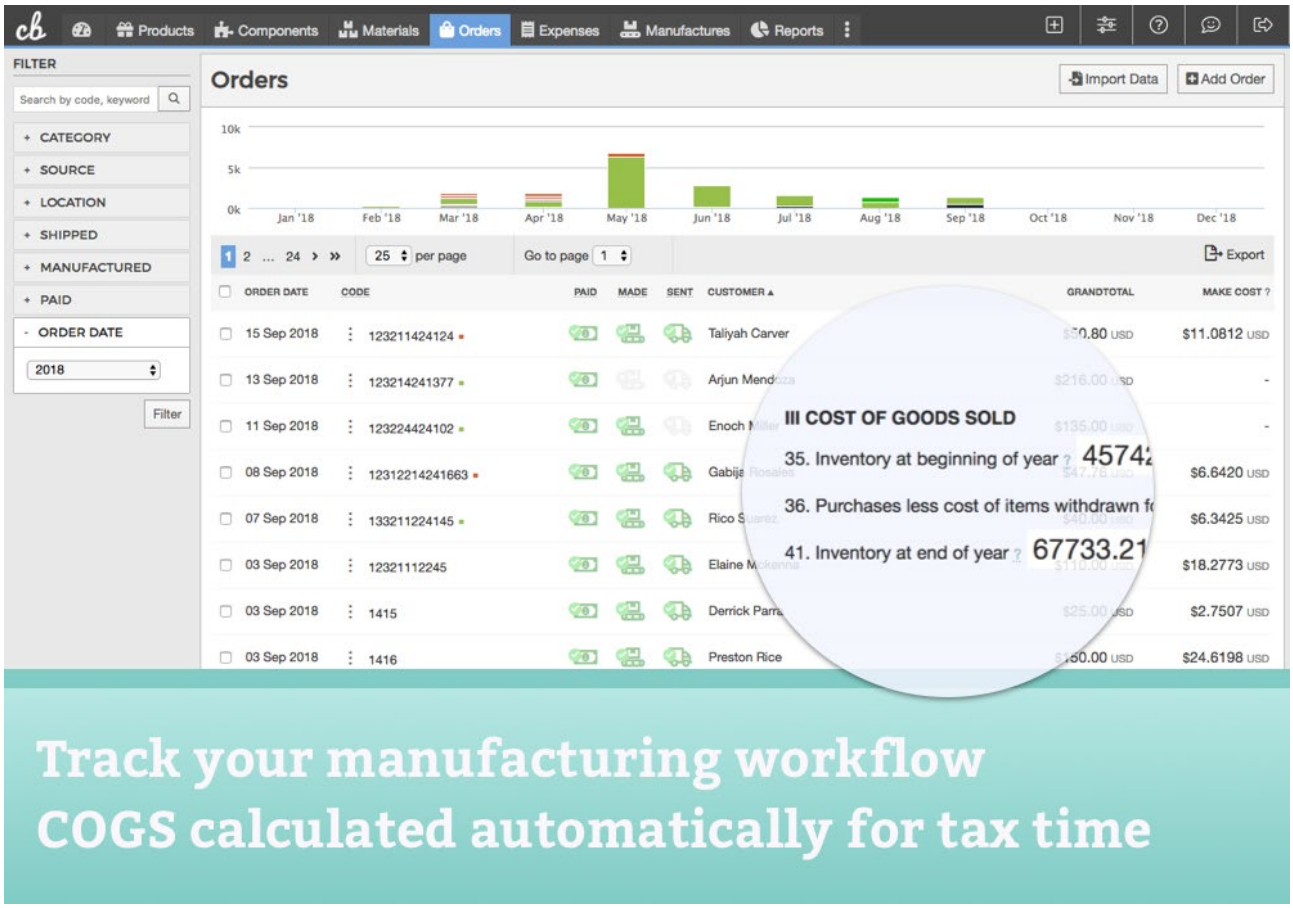

Рис. 4. Пример пользовательского интерфейса сервиса Craftybase

Сервис Marmalead

Marmalead [13] – это мощная система SEO аналитики, использующая в основе машинное обучение. Предлагает варианты поисковых запросов, анализ сезонности ключевых слов, прогнозирует, сколько уникальных посещений ожидается от ключевых покупателей в течение определенного периода. Один тарифный план – 19 долларов в месяц. Бесплатного периода нет. На рисунке 5 показан пример пользовательского интерфейса сервиса Marmalead.

Рис. 5. Пример пользовательского интерфейса сервиса Marmalead

Сервис Putler

Сервис Putler [14] объединяет данные из нескольких источников – торговые площадки, платежные системы, корзины покупок и Google Analytics – в одну систему. Предоставляет отчеты по прикрепленным магазинам и сайтам по различным направлениям: финансы, маркетинг, анализ клиентов, тренды, поисковая оптимизация, рассылки и т.д. Стоимость варьируется от 29 до 249 долларов в месяц. Бесплатный период – 14 дней. На рисунке 6 показан пример пользовательского интерфейса сервиса Putler.

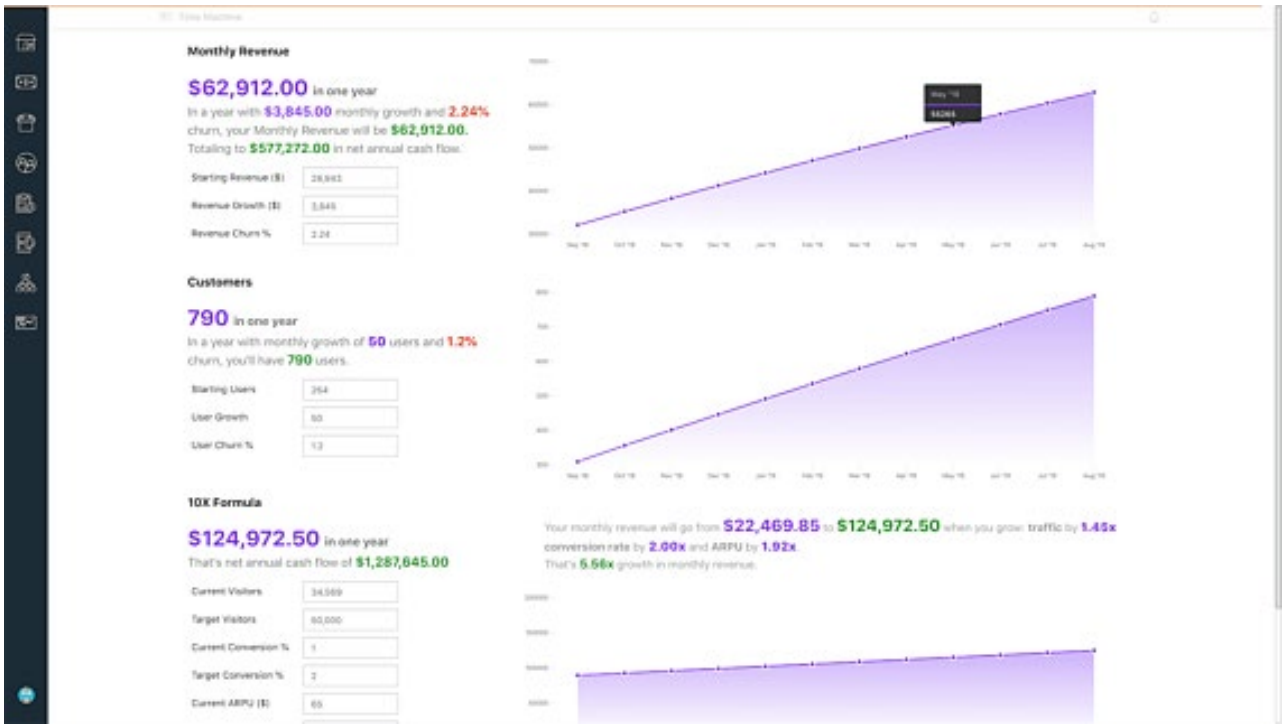

Рис. 6. Пример пользовательского интерфейса сервиса Putler

Putler – это единственный из всех сервисов, который предлагает более менее комплексное решение, остальные сосредоточены на одной, либо на нескольких сферах аналитики. Кроме того, все эти решения ориентированы на иностранных пользователей, и не имеют интерфейса на русском языке, что усложняет их использование.

Разработка специального приложения для анализа продаж магазина на площадке Etsy, ориентированного на русскоговорящего пользователя, упростит процесс ведения бизнеса для российского предпринимателя, сократив время, затрачиваемое на использование нескольких различных сервисов, уберет проблему необходимости перевода с иностранного языка на русский, и уменьшит расходы на вспомогательные программы.

1.4. ФУНКЦИОНАЛ РАЗРАБАТЫВАЕМОГО ПРИЛОЖЕНИЯ

Приложение предоставляет пользователю следующие функциональные возможности:

1. Получение сводной информации по магазину: дата первой продажи, общая сумма дохода за весь период в валюте магазина, общее количество проданных товаров, десять наиболее доходных товаров за

весь период активных продаж с указанием названия товара и суммы прибыли по этому товару, оформленных в виде таблицы.

2. Получение отчетности по выбранному году с указанием количества активных товаров, общей суммы дохода в валюте магазина, количества проданных товаров, десяти наиболее продаваемых товаров и десяти наиболее доходных товаров.
3. Выбор папки, содержащей csv файлы с данными отчетов магазина, по которым будет проведен анализ.
4. Получение списка товаров с возможностью выбора одного из трех вариантов (все товары / активные товары / неактивные товары) в виде таблицы, содержащей название товара, количество продаж данного товара, цену и общий доход по товару в валюте магазина, группу товара согласно ABC анализу и ссылку на анализ сезонности по товару.
5. Получение результатов анализа покупательской корзины в виде таблицы, содержащей основной товар и товар-компаньон.
6. Получение прогноза на период в двенадцать месяцев в виде таблицы с названием месяца и суммой предполагаемого дохода.
7. Получение результатов анализа сезонности отдельно по каждому товару в виде графика доходности и графика количества продаж.
8. Получение результатов ABC анализа доходности товаров в виде графика уровня доходности групп и таблицы, содержащей группу, количество товаров и процент товаров от общего числа.

1.5. ОПИСАНИЕ ДАННЫХ, ИСПОЛЬЗУЕМЫХ ДЛЯ АНАЛИЗА ПРОДАЖ МАГАЗИНА

Анализ продаж и разработка рекомендаций по оптимизации складских запасов проводится на основании отчетов магазина. Etsy предлагает владельцам магазинов несколько вариантов статистики для анализа. Данные предоставляются в формате .csv и доступны для скачивания на соответствующей странице пользовательского интерфейса.

Отчет EtsySoldOrderItems содержит данные по каждому проданному товару, такие как: дата продажи, номер заказа, данные о скидках, данные

покупателя, адрес доставки, стоимость товара. Полное описание структуры файла представлено в приложении 1. На рисунке 7 показан пример данных из отчета EtsySoldOrderItems2015.csv.

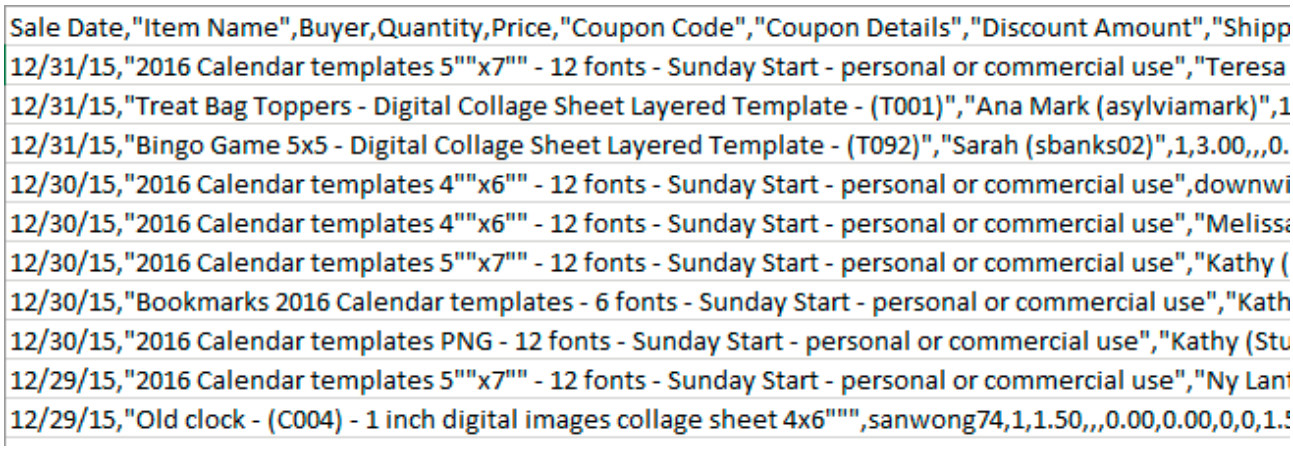

Рис. 7. Пример данных из отчета EtsySoldOrderItems2015.csv

Всего в отчете содержится 32 атрибута, из них для анализа используются 6:

1. Sales Date – дата продажи, тип атрибута – date, пример данных – «12/31/18»;
2. Item Name – название товара, тип атрибута – string, пример данных – «Treat Bag Toppers - Digital Collage Sheet Layered Template»;
3. Quantity – количество, тип атрибута – int, пример данных – «2»;
4. Price – цена, тип атрибута – float, пример данных – «2,50»;
5. Currency – валюта, тип атрибута – stringe, пример данных – «USD»;
6. Transaction ID – номер заказа, тип атрибута – int, пример данных – «1559656865».

Эти данные применяются для:

- анализа сезонности продаж конкретных товаров, их доходности;

- составления прогноза на период;

- анализа корзины.

Отчет EtsyListingsDownload содержит данные по каждому активному листингу: название, описание, цену, количество, ключевые слова, использованные материалы, ссылки на фотографии. Всего в отчете содержится от 8 до 17 атрибутов в зависимости от количества фотографий товара. Полное

описание структуры файла представлено в приложении 2. На рисунке 8 показан пример данных из отчета EtsyListingsDownload.csv.

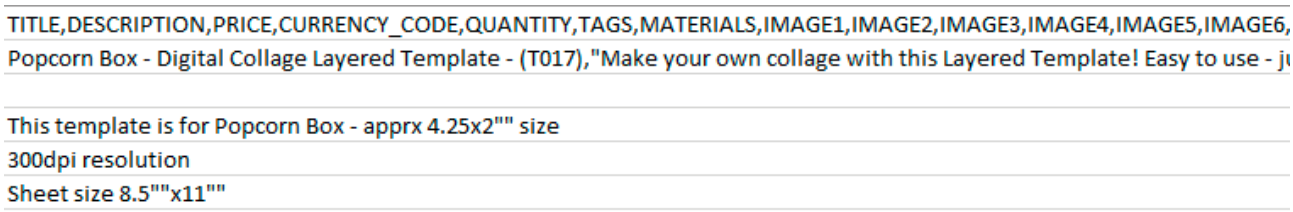

Рис. 8. Пример данных из отчета EtsyListingsDownload.csv

Так как этот отчет используется только для выбора активных товаров (то есть товаров, на данный момент находящихся в продаже), то для анализа используется только один атрибут: Title – название товара, тип атрибута – string, пример данных – Mini Candy Bar Wrappers - Digital Layered Template.

Отчет EtsySoldOrders содержит общие данные по заказам. В него включены такие атрибуты, как: дата заказа, номер заказа, сумма транзакции, данные о покупателе, тип оплаты заказа и так далее. Полное описание структуры файла представлено в приложении 3. На рисунке 9 показан пример данных из отчета EtsySoldOrders2015.csv.

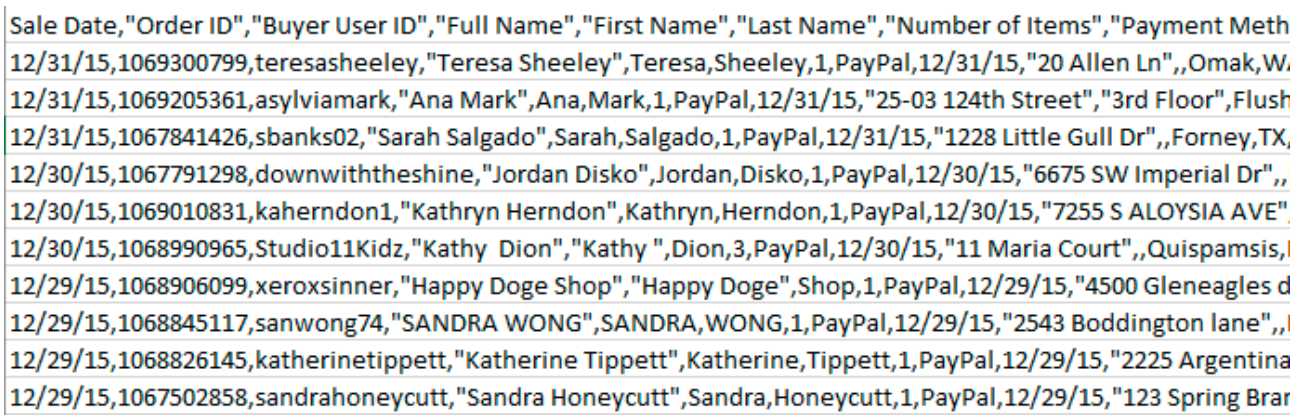

Рис. 9. Пример данных из отчета EtsySoldOrders2015.csv

Всего в отчете содержится 33 атрибута, из них для анализа используются 5:

1. Sale Date – дата продажи, тип атрибута – date, пример данных – «01.01.2020»;
2. Order ID – номер заказа, тип атрибута – int, пример данных – «1566937513».
3. Number of Items – количество товаров в заказе, тип атрибута – int, пример данных – «2»;

4. Currency – валюта, тип атрибута – stringe, пример данных – «USD»;
5. Order Value – сумма заказа, тип атрибута – float, пример данных – «5,00»

Данные из этого отчета применяются для:

- анализа сезонности продаж конкретных товаров, их доходности;
- анализа корзины.

ГЛАВА 2. ОСНОВНЫЕ ПРИМЕНЯЕМЫЕ АЛГОРИТМЫ

2.1. ЗАДАЧА ПОИСКА АССОЦИАТИВНЫХ ПРАВИЛ

Обучение на ассоциативных правилах (далее Associations rules learning – ARL) – это метод машинного обучения, позволяющий находить отношения между переменными в больших объемах данных. Это часто применяемый метод поиска взаимосвязей (ассоциаций) в данных.

Впервые задача поиска ассоциативных правил была предложена для нахождения типичных шаблонов покупок, совершаемых в супермаркетах, поэтому иногда ее еще называют анализом рыночной корзины (market basket analysis). В общем виде этот метод можно описать как «Кто купил x_1, также купил x_2». В основе метода лежит анализ транзакций, каждая из которых содержит свой уникальный itemset из набора items.

Основные используемые понятия:

- Множество объектов (itemset):

$$X \subseteq I = \{x_1, x_2, \dots, x_n\} \quad (1)$$

- Множество идентификаторов транзакций (tIDset):

$$T = \{t_1, t_2, \dots, t_n\} \quad (2)$$

- Множество транзакций (transactions):

$$\{(t, X): t \in T, X \in I\} \quad (3)$$

С помощью алгоритмов ARL внутри одной транзакции находятся «правила» – совпадения items, которые потом сортируются по их силе. Ассоциативное правило – «из события A следует событие B», то есть если в itemset есть x_1, то есть и x_2.

Данный алгоритм применяется на основе данных, найденных в процессе анализа информации об ассортименте товаров в чеке.

Анализ рыночной корзины позволяет определить взаимосвязь между товарами, купленными в разных транзакциях, что дает возможность спланировать дальнейшие действия для увеличения прибыли. Например, если товары x_1 и x_2 покупаются вместе чаще, можно предпринять следующее:

- Товары x_1 и x_2 могут быть размещены вместе, чтобы, когда покупатель выбирает один из продуктов, ему проще будет найти другой продукт.
- Люди, которые покупают один из продуктов, могут быть нацелены на покупку другого с помощью рекламной кампании.
- При покупке товаров x_1 и x_2 вместе может предложена скидка.
- Товары x_1 и x_2 могут продаваться комплектом.

Базовые понятия в ARL [15]

Support (поддержка):

$$supp(X) = \frac{\{t \in T; X \in t\}}{|T|} \quad (4)$$

где t – транзакция, X – набор элементов, содержащий в себе x, а T – количество транзакций. То есть в общем виде support – это показатель того, как часто встречается данный набор элементов в транзакциях, взятых для анализа. Если рассматривать вариант, когда в одном itemset встречаются item x_1 и item x_2, то нужно посчитать, во скольких транзакциях встречается эта пара.

$$supp(x_1 \cup x_2) = \frac{\sigma(x_1 \cup x_2)}{|T|} \quad (5)$$

где σ – это количество транзакций, содержащих x_1 и x_2.

Confidence (достоверность)

Достоверность – это показатель того, как часто полученное правило работает для всего набора данных. Вычисляется по следующей формуле:

$$conf(x_1 \cup x_2) = \frac{supp(x_1 \cup x_2)}{supp(x_1)} \quad (6)$$

Таким образом вычисляется, во скольких транзакциях, содержащих x_1, также имеется x_2.

Lift (лифт, подъемная сила)

Lift показывает, насколько элементы x зависят друг от друга:

$$lift(x_1 \cup x_2) = \frac{conf(x_1 \cup x_2)}{supp(x_2)} \quad (7)$$

Например, требуется понять зависимость x_1 и x_2. Для этого считаем c*onfidence* правила $x_1 + x_2$ и делим его на support x_2. Если lift равен единице, можно утверждать, что элементы независимы и не имеют правил совместной покупки. Если lift больше единицы, то «сила» правила, это величина, на которую lift больше единицы. Если lift меньше единицы, то это показывает, что правило основания x_2 негативно влияет на правило x_1.

Алгоритм поиска ассоциативных правил предназначен для нахождения всех существующих закономерностей для массива транзакций. При этом support и confidence этих правил должны быть выше некоторых заранее определенных уровней, которые называются соответственно minsupport (минимальная поддержка) и minconfidence (минимальная достоверность).

Значения для параметров minsupport и minconfidence выбираются таким образом, чтобы количество сгенерированных правил не было слишком большим, или слишком маленьким. Если minsupport имеет высокое значение, то алгоритмы будут находить только хорошо известные и очевидные зависимости, и анализ не будет иметь ценности. При этом, низкое значение параметра minsupport приведет к генерации большого числа правил, которые могут оказаться статистически необоснованными.

Есть несколько часто применяемых алгоритмов, которые позволяют находить ассоциативные правила в наборе данных согласно перечисленным выше понятиям, например, FP-growth [16], ECLAT [17], Apriori [18] и другие. Для реализации использовался алгоритм Apriori.

Процесс работы алгоритма Apriori можно разбить на несколько этапов:

- Установить минимальное значение support и confidence;
- Извлечь все подмножества, имеющие более высокое значение поддержки, чем минимальный порог;
- Выбрать все правила из подмножеств со значением достоверности выше минимального порога;
- Установить правила в порядке убывания лифта.

Псевдокод алгоритма Apriori:

ВХОД: Датасет D, содержащий список транзакций, и σ – задаваемый пользователем порог support

ВЫХОД: Список itemsets F(D, σ)

ПОДХОД:

1. $C_1 \leftarrow [\{i\}|i\ in\ J]$
2. $k \leftarrow 1$
3. $while\ C_k \neq 1\ do$:
4. #Считаем все $support$ для всех кандидатов

   $for\ all$ транзакций $(tid, I) \in D\ do$:

   $for\ all$ кандидатов $X \in C_k\ do$:

   $if\ X \in I$:

   $X.support++$
5. #Вытаскиваем все частые $itemsets$ для всех кандидатов

   $F_k = \{X|X.support > \sigma\}$
6. #Генерируем новых кандидатов

   $\forall\ X, Y \in F_i, X_{[i]} = Y_{[i]}\ for\ 1 \leq i \leq k-1\ and\ X_{[k]} \leq Y_{[k]}\ do$:

   $I = X \cup \{Y_{|k|}\}$

   $if\ \forall\ J\ \subseteq I, |J| = k: J \in F_k\ then$

   $C_k + 1 \in C_k + 1 \cup I$

   $k++$

Таким образом, первом шаге алгоритм Apriori ищет все наборы товаров с одним элементом, удовлетворяющие заданному значению параметра support, на втором шаге составляет из найденных пары по принципу иерархической монотонности, то есть если x_1 встречается часто и x_2 встречается часто, то и x_1+x_2 встречается часто.

Поиск всех часто встречающихся элементов требует больших вычислительных и временных ресурсов. Для снижения размерности пространства поиска алгоритм Apriori использует так называемое свойство анти-монотонности. С увеличением размера набора элементов значение параметра

support уменьшается, либо не меняется. Суть свойства анти-монотонности в том, что значение параметра support для любого набора элементов не может превышать минимального значения support любого из подмножеств этого набора элементов. На рисунке 10 представлена визуализация свойства анти-монотонности.

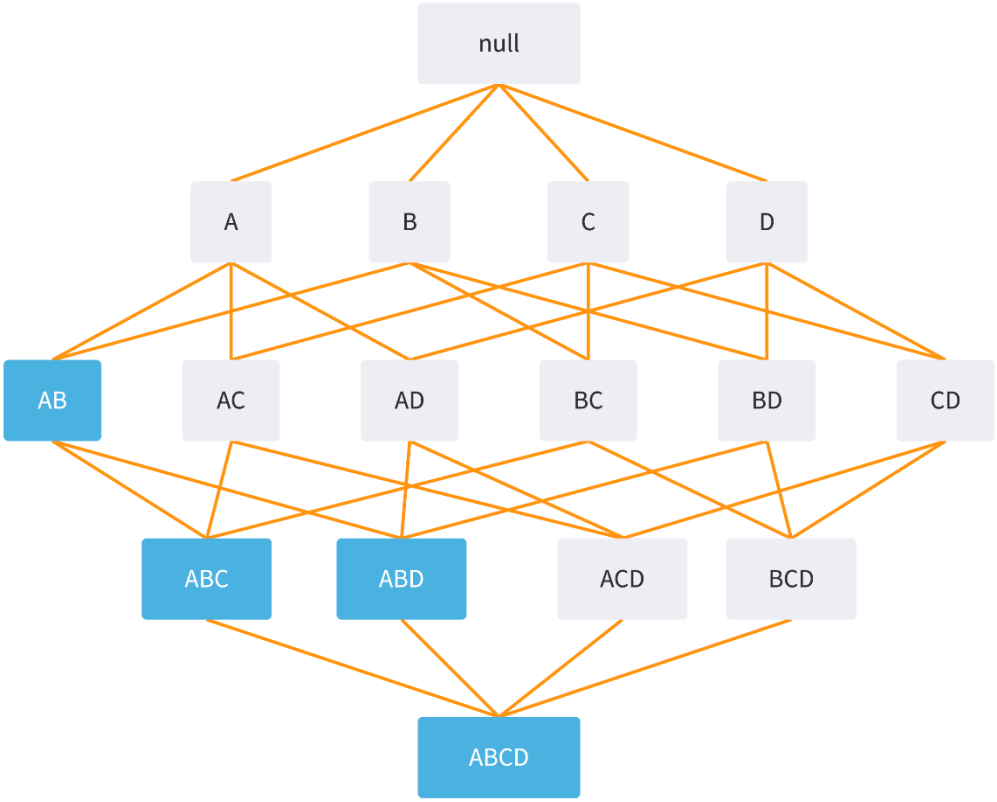

Рис. 10. Реализация алгоритма Apriori

Допустим, набор элементов AB имеет значение параметра support ниже заданного порога, и таким образом не является часто встречающимся. Из этого следует, что все супермножества, содержащие набор элементов AB, также не является часто встречающимися.

2.2. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ АНАЛИЗА ДОХОДНОСТИ (ABC)

Одним из популярных среди владельцев бизнеса методов исследования данных о продажах магазина является анализ структуры товарооборота, или ABC-анализ [19]. Цель этого анализа – выявить долю того или иного продукта в общем объеме продаж. В основе анализа лежит “Принцип Парето” [20]: 20% ресурсов приносят 80% прибыли. Этот вид анализа позволяет разделить товары

на группы, и определить, какие приносят компании основную прибыль, а какие являются убыточными.

На основе данных о прибыли выделяются три группы товаров:

- группа А – товары, приносящие основной доход, – от 0 до 80%;
- группа В – товары, пользующиеся спросом, но приносящие от до 15% дохода;
- группа С – товары этой группы приносят наименьший доход, как правило около 5%, их продажа не является прибыльной.

Процентное соотношение 80 / 15 / 5 – это классическая схема разделения на группы, однако на практике проценты могут быть любыми в зависимости от ассортимента товаров, специфики магазина, ценовой политики, целевой аудитории.

2.3. ОБЗОР МЕТОДОВ ПРОГНОЗИРОВАНИЯ

Одним из самых важных условий принятия верных управленческих решений в производственных и торговых организациях является прогнозирование спроса. Этот параметр влияет на процессы закупки, производства, планирования рабочего времени, логистику и так далее.

Адаптивные методы являются одними из перспективных направлений развития прогнозных моделей. С помощью этих методов можно создавать модели, обладающие возможностью изменяться и реагировать на меняющиеся условия [21]. Адаптивные методы исследуют и учитывают различную информацию данных временных рядов, что позволяет эффективно применять их для прогнозирования неустойчивых временных рядов со склонностью к изменению. Для работы с уровнями временных рядов используется система весов.

Многие из базовых методов прогнозирования относятся скорее к отдельным приемам или процедурам, другие являются пакетами методов и отличаются друг от друга количеством частных приемов и/или последовательностью их применения [22].

В зависимости от объема формулировки все методы прогнозирования делятся на интуитивные и формальные. Интуитивное предсказание

используется, когда объект настолько сложен, что практически невозможно игнорировать влияние многих факторов. Формальные прогнозы строятся с использованием компьютерных математических методов, и вы можете получить наиболее надежные данные за меньшее время.

По степени формализации методы прогнозирования можно разделить на интуитивные и формализованные. Интуитивные методы прогнозирования применяются, когда нет полной информации об исследуемом объекте, нет возможности формально описать причинно-следственные связи, очень сложно учесть влияние многих факторов. Суть методов состоит в получении прогнозных оценок развития состояния объекта во времени путем проведения интуитивно-логического анализа проблемы в сочетании с количественной оценкой суждений и формальном описании результатов анализа. Так как эта группа методов предполагает использование мнений людей-экспертов, их еще называют методами экспертных оценок.

Формализованные методы прогнозирования в свою очередь используют формальные средства математической теории для исследования и предсказания поведения объекта во времени. Применение математических методов исследования для прогноза позволяет значительно сократить время, повысить точность результатов прогнозирования и достоверность прогнозов, упростить обработку информации в целом.

Модель скользящего среднего (MA)

Один из хорошо известных методов сглаживания временных рядов – это метод скользящего среднего [23]. С помощью этого метода можно исключить случайные колебания ряда и получить значения в зависимости от влияния ключевых факторов.

Способ сглаживания временного ряда при помощи скользящего среднего основан на том, что внутри выбранного интервала времени первоначальные значения ряда заменяются на среднее арифметическое. В процессе период сдвигается на один интервал и расчет среднего повторяется. Интервал всегда остается одним и тем же, а получаемое среднее арифметическое значение

относится к средине выбранного временного периода. В результате в каждом интервале сглаживания среднее представляет собой новую точку.

Модель скользящего среднего q-го порядка – это модель временного ряда вида:

$$x_t = \sum_{j=0}^{q} b_j \varepsilon_{t-j} \quad (8)$$

где x – члены ряда, t – момент времени, q – порядок скользящего среднего, b – параметры модели, ε – белый шум [24].

Метод скользящих средних подходит как для равномерно изменяющегося временного ряда, так и для данных с неявно выраженным трендом. Скользящее окно широко применяется для предобработки данных в прогнозировании и других видах анализа, поскольку позволяет исключить влияние случайной составляющей. Однако при сильных скачках ближе к «будущему» он может давать сбои. Недостаток метода состоит в его локальности – он не реагирует на данные в целом и не может прогнозировать резкое изменение поведения данных, опираясь лишь на ближайшие точки.

Линейный метод наименьших квадратов

Статистически линейная регрессия [25] – это линейный метод, используемый для установления связи между зависимой переменной и одной или несколькими независимыми переменными. В случае одной независимой переменной это называется простой линейной регрессией, если переменных несколько, то это множественная линейная регрессия.

Уравнение линейной регрессии имеет вид: $y = \alpha x + b$, где α и b – коэффициенты линейного уравнения, x – независимая переменная, а y – зависимая переменная.

Метод наименьших квадратов (МНК) заключается в том, что наилучшим приближением к реальным данным будет минимизация суммы квадратов разностей между фактическим уровнем и теоретическим:

$$\sum_{i=1}^{n}(y_i - \alpha x_i - b)^2 \rightarrow min \quad (9)$$

Линейный метод наименьших квадратов используется для решения задач сглаживания данных, экстраполяции и интерполяции.

Модель Брауна

В модели Брауна [26], также известной как метод экспоненциального сглаживания, динамический ряд сглаживается с помощью взвешенной скользящей средней, в которой значения подчиняются экспоненциальному закону. Идея метода заключается в том, что прогнозное значение $\bar{y}_{t+1}$ определяется через предыдущее спрогнозированное значение $\bar{y}_t$, но скорректированное с некоторым коэффициентом на величину отклонения факта y_t от прогноза: $\bar{y}_{t+1} = \bar{y}_t + \alpha(y_t - \bar{y}_t)$. Метод экспоненциального сглаживания аналогичен методу скользящего среднего. У них есть общий главный принцип – каждая точка зависит от значений соседних с некоторыми весами. Главные отличия метода Брауна – начальная точка не затрагивается и остается неизменной, по мере удаления от начальных данных веса экспоненциально стремятся к нулю, кроме того, данные сглаживаются не в центре, а к ближайшему предыдущему значению.

Метод Брауна подходит для данных с зависимостью от предыдущих значений и без сильных амплитудных колебаний.

Рассматриваемый метод прогнозирования является достаточно эффективным и надежным. Но он дает возможность спрогнозировать процесс только в краткосрочном периоде, т.е. всего лишь на 1–2 года вперед.

Авторегрессионная модель (AR)

С помощью AR-моделей [27] моделируется сезонность временного ряда. В отличие от MA-модели, где величина временного ряда в момент *t* зависит от авторегрессивных белых шумов, в этой модели величина временного ряда в момент *t* (настоящий момент) зависит от предыдущих значений этого же ряда. То есть, данная модель строится из предположения о том, что каждый член временного ряда образуется при помощи p предыдущих членов:

$$y_t = c + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t \quad (10)$$

где y – члены ряда, c = константа, a – коэффициенты авторегрессии, ε_t – белый шум. Таким образом, в модели уже заложен прогноз на p шагов вперед при имеющихся начальных значениях y_i. Наиболее рациональным способом определения параметров уравнения авторегрессии является использование метода наименьших квадратов. Однако главная задача сводится к поиску порядка авторегрессии p.

Применение авторегрессионных моделей основано на предварительном анализе, когда известно, что изучаемый процесс в значительной степени зависит от его развития в предыдущие периоды. Область применения авторегрессионной модели ограничена, поскольку помимо сезонных изменений она ничего не описывает. Она редко используется в чистом виде. В большинстве случаев применяется более гибкая модель, включающая скользящее окно.

Модель авторегрессии – скользящего среднего (ARMA)

Модель ARMA [28] представляет собой сумму авторегрессии – запаздывания данных – и скользящего окна:

$$y_t = c + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t - \sum_{j=1}^{q} b_j \varepsilon_{t-j} \quad (11)$$

Эта модель хорошо подходит для прогнозирования данных, стабильно сохраняющих свою динамику на всем рассматриваемом временном периоде, то есть при предположении, что временной ряд стабилен, его свойства не меняются во времени.

Модель Бокса-Дженкинса (ARIMA)[29]

Так называемая интегрированная модель авторегрессии – скользящего среднего, расширенная версия модели ARMA. ARIMA строится на предположении о том, что данные имеют авторегрессию, шумовой эффект и интеграцию.

$$\Delta^d x_t = c + \sum_{i=1}^{p} \Delta^d x_{t-i} + \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j} \quad (12)$$

Интеграция данных подразумевает наличие стабильной разности некоторого порядка. То есть $\Delta^d x_t = x_{t-d} - x_t \ \forall t$ сохраняет свой вид и поведение. Модель достаточно часто применяется для прогнозирования временных рядов и имеет много улучшений и вариаций. Метод считается достаточно точным, чтобы получать как краткосрочные, так и долгосрочные прогнозы, которые не требуют отдельного оценивания. Однако для построения модели типа ARIMA требуется больший набор данных и всеобъемлющий их анализ.

Модель Бокса-Дженкинса может быть применима к большому количеству типов данных, однако требует большей выборки (около 40 точек для хорошего прогноза) и тщательного исследования поведения временного ряда. Недостаток заключается в том, что построение удовлетворительной модели ARIMA требует больших затрат ресурсов и времени.

Метод Холта-Винтерса

Экспоненциальное сглаживание Холта-Винтерса [30] используется для прогнозирования данных временных рядов, которые демонстрируют как наличие тренда, так и сезонные колебания. Метод Холта-Винтерса состоит из следующих четырех методов прогнозирования, наложенных друг на друга:

- Взвещенная скользящая средняя;
- Экспоненциальное сглаживание;
- Экспоненциальное сглаживание Холта;
- Экспоненциальное сглаживание Холта-Винтерса.

1. Взвещенная скользящая средняя – это средняя из n чисел, где каждому числу присваивается определенный вес, а знаменатель равен сумме этих n весов. Веса часто назначаются согласно какой-либо весовой функции. Обычными весовыми функциями являются логарифмические, линейные, квадратичные, кубические и экспоненциальные. Усреднение как метод прогнозирования временных рядов имеет свойство сглаживать вариации исторических значений при расчете прогноза. Выбирая подходящую

весовую функцию, прогнозист определяет, каким историческим значениям следует уделить особое внимание при вычислении будущих значений временного ряда.

2. Метод экспоненциального сглаживания прогнозирует следующее значение, используя средневзвешенное значение всех предыдущих значений, где веса экспоненциально убывают от самого последнего к самому старому историческому значению. При использовании этого метода предполагается, что недавние значения временного ряда намного важнее, чем более старые значения. Метод экспоненциального сглаживания нельзя использовать, если данные имеют тренд и / или сезонные колебания.
3. Экспоненциальное сглаживание Холта устраняет один из двух недостатков простого метода экспоненциального сглаживания. Этот метод можно использовать для прогнозирования данных временных рядов, имеющих тенденцию. Но при наличии сезонных колебаний метод показывает неточные результаты.
4. Метод Холта-Винтерса изменяет методику экспоненциального сглаживания Холта таким образом, что его становится возможным использовать как при наличии тренда, так и при наличии сезонности.

Пусть задан временной ряд $y_0 \ldots y_t$, $y_i \in R$. Необходимо решить задачу прогнозирования временного ряда:

$$\begin{cases} \bar{y}_{t+d} = (a_t + kr_t)\Theta_{t+k-s} \\ a_t = \alpha \dfrac{y_t}{\Theta_{t-s}} + (1-\alpha)(a_{t-1} + r_{t-1}) \\ r_t = \gamma(a_t - a_{t-1}) + (1-\gamma)r_{t-1} \\ \Theta_t = \beta \dfrac{y_t}{a_t} + (1-\beta)\Theta_{t-s} \end{cases} \quad (13)$$

где s – период сезонности, Θ_i, $i \in [0, s-1]$ – сезонный профиль, r_t – параметр тренда, a_t – параметр прогноза, очищенный от влияния тренда и сезонности. Соответственно коэффициент $\alpha \in [0, 1]$ указывает, как сильно величины зависят от предыдущих значений, $\beta \in [0, 1]$ показывает значимость сезонности, а $\gamma \in [0,$

1] – есть ли у данных ярко выраженный тренд. Оптимальные параметры α, β, γ предлагается находить экспериментальным путем.

Метод Холта-Винтерса универсален для перечисленных выше особенностей, однако для данных нужно строгое определение сезонности. Этот метод может применяться:

- при стратегическом планировании: построение основной тенденции развития (тренда) дает возможность учитывать восходящую или нисходящую динамику исследуемого явления;

- при оперативном и тактическом планировании: выявленная сезонная составляющая позволяет отметить неравномерность распределения объемов по годам по отношению к данной динамике.

Экспоненциальное сглаживание учитывает внутренние спады и подъемы в ряде динамики. Его можно использовать при выявлении крупных спадов и подъемов заблаговременно (при применении тактического планирования) и быть к ним готовым. Таким образом, метод имеет достаточно большую сферу применения. Данный метод основан на использовании большого объема статистических данных, что не всегда может быть актуально. Метод Холта-Винтерса может применяться при комбинированном прогнозировании одновременно с экспертными методами прогнозирования.

С учетом специфики продаж интернет-магазина наиболее целесообразным для обработки статистических данных и прогнозирования на дальнесрочные, среднесрочные и краткосрочные периоды является применение модели Холта-Винтерса, которая учитывает экспоненциальный тренд (тенденция изменения показателей временного ряда) и аддитивную сезонность (периодические колебания, наблюдаемые на временных рядах).

ГЛАВА 3. РАЗРАБОТКА ПРОГРАММНОГО ПРОДУКТА

3.1. ВЫБОР СРЕДЫ РАЗРАБОТКИ ПРИЛОЖЕНИЯ

Реализация описанных моделей интеллектуального анализа данных для исследования продаж интернет-магазина на площадке Etsy выполнена в виде десктопного приложения. В качестве источника данных принимающее файлы отчетов магазина Etsy. Данные предоставляются в формате .csv и доступны для скачивания в соответствующей секции аккаунта Etsy. Программа написана на python 3.8. GUI сделан при помощи программы Qt Designer и библиотеки PyQt5.

Qt – кроссплатформенный фреймворк для разработки программного обеспечения на языке программирования C++. Qt дает возможность использовать созданное с его помощью программное обеспечение во многих современных операционных системах без изменения исходного кода при помощи простой компиляции программы для каждой системы. Содержит все основные классы, которые могут быть необходимы для разработки прикладного программного обеспечения, включая классы для работы с базами данных, сетью, элементы графического дизайна и так далее. Qt – это объектно-ориентированный фреймворк, который поддерживает технику компонентного программирования.

PyQt [31] – набор расширений Qt для языка программирования python, реализованный в виде библиотеки python. PyQt разработан британской компанией Riverbank Computing. Работает на всех платформах, поддерживаемых Qt: Linux и другие UNIX-подобные ОС, Mac OS X и Windows. Существует 2 версии: PyQt5, поддерживающий Qt 5, и PyQt4, поддерживающий Qt 4.

PyQt практически полностью реализует возможности Qt. Это более 600 классов, более 6000 функций и методов, включая существующий набор виджетов графического интерфейса; стили виджетов; доступ к базам данных с помощью SQL (ODBC, MySQL, PostgreSQL, Oracle); QScintilla, основанный на Scintilla виджет текстового редактора; парсер XML; поддержку SVG; интеграцию с WebKit, движком рендеринга HTML; поддержку воспроизведения видео и аудио.

PyQt также включает в себя Qt Designer (Qt Creator) – инструмент для проектирования и создания графических пользовательских интерфейсов (GUI) из компонентов Qt. Позволяет создавать и настраивать свои виджеты или диалоги в режиме "что видишь, то и получишь" (what-you-see-is-what-you-get, WYSIWYG). Приложение pyuic генерирует код на python из файлов, созданных в Qt Designer. Это делает PyQt очень полезным инструментом для быстрого прототипирования. Также в Qt Designer присутствует возможность добавлять новые графические элементы управления, написанные на python.

Для работы с таблицами использовались библиотеки pandas и numpy. Для построения графиков использовалась библиотека matplotlib.pyplot. Для работы с операционной системой использовался модуль os. Библиотека datetime использовалась для работы с датами. Для реализации алгоритма apriori была использована библиотека apyori, для реализации алгоритма Холта-Винтерса был использован модуль statsmodels.

3. 2. СТРУКТУРА ПРОГРАММЫ

Приложение запускается в оконном режиме, и не имеет функции полноэкранного режима. Интерфейс пользователя включает в себя восемь разделов, каждый из которых предоставляет либо статистическую, либо прогнозную информацию. Доступ к разделам организован в виде одноуровневого меню, содержащего восемь кнопок, которые позволяют перейти в соответствующий раздел.

Основные файлы программы:

- *main.py* – исполняемый файл программы;
- *design.py*, *mplwidget.py*, *design_dev.py* – файлы, автоматически сгенерированные при помощи программы Qt Designer в процессе проектирования пользовательского интерейса;
- *forecast.py*, *info_page.py*, *items_page.py*, *main_page.py*, *ABC_analysis.py*, *basket.py* – файлы, содержащие функции для проведения статистических и прогнозных расчетов.

Файл *main.py* – исполняемый файл программы, который содержит все методы, запускающие отдельные функции, которые отвечают за выполнение необходимых процессов. На рисунке 11 представлен пример кода метода, который обращается к функции *basket_info*, находящейся в файле *basket.py*, и отвечающей за анализ потребительской корзины.

```
def basket_loader(self):
    global ys
    basket.bascket_info(ys, self.path)
    bascket_df = pd.read_csv(self.path + '/basket_table_' + '_'
                             .join([str(y) for y in ys]) + '.csv')
    del bascket_df['Unnamed: 0']
    bascket_df = bascket_df[['ОСНОВНОЙ ТОВАР', 'ТОВАР_КОМПАНЬОН']]
    set_table(self.tableWidget_5, bascket_df)
    return self.stackedWidget.setCurrentWidget(self.basket_page)
```

Рис. 11. Пример кода метода, переводящего на страницу анализа корзины

В приложении 4 приведены базовые функции взаимодействия с пользовательским интерфейсом.

Приложение не требует отдельной базы для хранения данных. Промежуточные данные хранятся в виде файлов в формате .csv и .txt в выбранной пользователем папке.

3.3. РАЗРАБОТКА ИНТЕРФЕЙСА ДЛЯ ПОЛУЧЕНИЯ СВОДНОЙ ИНФОРМАЦИИ ПО МАГАЗИНУ

Для отображения сводной информации по магазину используются данные отчетов EtsySoldOrderItems за все года активности магазина. В этом разделе программы предоставляется следующая информация:

- дата первой продажи;
- количество проданных товаров;
- общая сумма дохода в валюте магазина;
- десять наиболее прибыльных товаров за весь период активных продаж.

Дата первой продажи обычно относится к ключевым датам магазина, таким, как основные праздники, официально определенные недели распродаж, значимые для владельца даты. К таким датам приурочивают дни скидок или

составляют специальные акционные предложения для привлечения покупателей.

Количество проданных товаров и общая сумма дохода – это общая статистическая информация, которая может потребоваться для составления планов или отметки о достигнутых целях.

Десять наиболее прибыльных товаров показывают, на какие товары стоит обратить внимание для повышения уровня продаж. Эти данные представлены в виде таблицы с указанием названия товара и суммы дохода в валюте магазина за весь период продаж.

На рисунке 12 представлен интерфейс пользователя, позволяющий просмотреть сводную информацию по магазину.

Рис. 12. Интерфейс «Сводная информация по магазину»

3.4. РАЗРАБОТКА ИНТЕРФЕЙСА ДЛЯ ПОЛУЧЕНИЯ СТАТИСТИЧЕСКИХ ОТЧЕТОВ О ПРОДАЖАХ ЗА ОТДЕЛЬНЫЙ ГОД

Получение основной статистической информации, такой как сумма дохода, количество продаж и так далее, является важным для отслеживания этапов планирования и работы магазина.

Интерфейс «Отчет по годам» предоставляет следующую информацию:

- Количество активных товаров в выбранный период;
- Количество продаж за выбранный период;
- Сумма выручки за выбранный период;

- Десять наиболее продаваемых товаров за выбранный период;
- Десять наиболее доходных товаров за выбранный период.

Информация предоставляется за период в двенадцать месяцев с января по декабрь. Выбор года для просмотра информации осуществляется при помощи выпадающего меню.

Метод *info* подготавливает данные для страницы с отчетом по году.

```
def info(y, path):
    df = pd.read_csv(path + f'/EtsySoldOrderItems{y}.csv')
    active_items = len(pd.unique(df['Item Name']))
    sold_count = df['Quantity'].sum()
    sold_sum = df['Item Total'].sum()
    df = df.groupby('Item Name')[['Quantity', 'Item Total']].sum()
    df_1 = df.sort_values(by='Quantity', ascending=False)
    ten_count = pd.DataFrame(data=np.array([df_1.index]).T,
columns=['Item Name']).iloc[:10]
    df_2 = df.sort_values(by='Item Total', ascending=False)
    ten_sold = pd.DataFrame(data=np.array([df_2.index]).T, columns=['Item
Name']).iloc[:10]
    return active_items, sold_count, sold_sum, ten_sold, ten_count
```

Этот метод содержится в файле info_page.py и вызывается при помощи метода *info_page_loader* в исполняемом файле.

На рисунке 13 представлен интерфейс пользователя, позволяющий просмотреть отчет по годам.

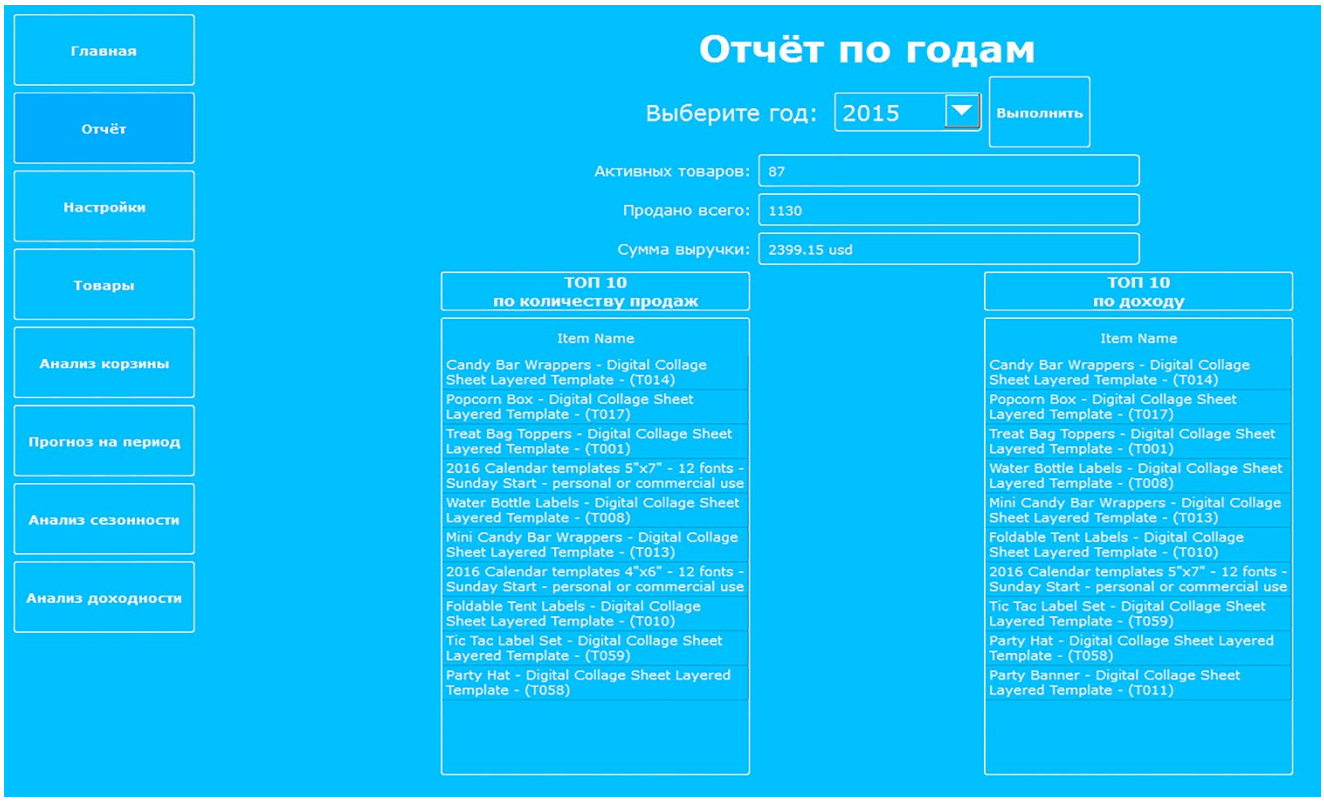

Рис. 13. Интерфейс «Отчет по годам»

3.5. РАЗРАБОТКА ИНТЕРФЕЙСА ДЛЯ ЗАГРУЗКИ ОТЧЕТОВ ETSY

Выгрузка статистических отчетов с сайта Etsy производится непосредственно пользователем. Для корректной работы разрабатываемого приложения файлы данных в формате csv необходимо поместить в отдельную папку и указать путь к этой папке. Для выполнения этой задачи был разработан интерфейс «Настройки». При помощи этого интерфейса пользователь может выбрать папку, куда предварительно были загружены файлы отчетов Etsy.

На рисунке 14 представлен интерфейс выбора папки для загрузки данных.

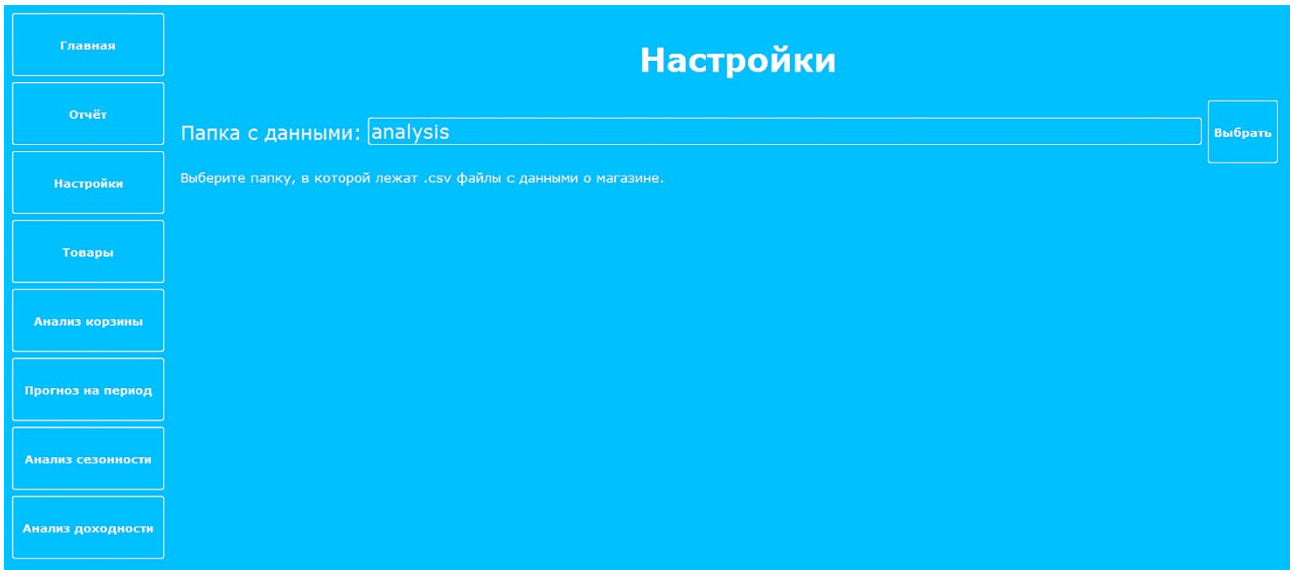

Рис. 14. Реализация интерфейса для загрузки отчетов Etsy

Интерфейс был разработан с использованием библиотеки os, которая предназначена для работы с директориями и командной строкой.

3.6. РАЗРАБОТКА ИНТЕРФЕЙСА ДЛЯ ПОЛУЧЕНИЯ СПИСКА ТОВАРОВ МАГАЗИНА

Данный интерфейс позволяет посмотреть либо все когда-либо продаваемые товары магазина, либо товары, сгруппированные по принципу «активные/неактивные». «Активные» товары – те, которые доступны для покупки в данные момент. «Неактивные» – товары, которые либо продавались, либо только готовятся к продаже, но не выставлены на продажу в настоящее время. Выбор принципа группировки товаров осуществляется при помощи выпадающего меню.

Данные представлены в виде таблицы, содержащей:

- Название товара;
- Количество проданного товара;
- Сумму дохода, полученную после продаж этого товара;
- Цену товара;
- Группу товара согласно ABC анализу;
- Ссылку на результат анализа сезонности этого товара.

Метод *set_item_page* визуализирует результат формирования списка товаров магазина:

```
def set_item_page(self):
        global ys
        ABC_analysis.abc(ys, self.path)
        state = self.comboBox.currentText()
        all_items, active_items, disactive_items =
items_page.items_info(ys, self.path)
        if state == 'Все':
            self.items_set_table(self.tableWidget, all_items)
        elif state == 'Активные':
            self.items_set_table(self.tableWidget, active_items)
        else:
            self.items_set_table(self.tableWidget, disactive_items)
        return self.stackedWidget.setCurrentWidget(self.items_page)
```

Так как интерфейс предполагает возможность просмотра анализа сезонности по отдельным товарам, был дополнительно разработан метод, позволяющий перейти на результаты анализа сезонности:

```
def items_seasonal(self, item):
        self.flag_items = 1
        self.items_item = item
        self.season_subpage_loader()
        self.flag_items = 0
        return
```

На рисунке 15 представлен интерфейс пользователя, позволяющий просмотреть список товаров магазина.

Рис. 15. Реализация интерфейса для получения списка товаров магазина

3.7. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ЗАДАЧИ АНАЛИЗА ПОТРЕБИТЕЛЬСКОЙ КОРЗИНЫ

Модуль Apyori – это простая реализация алгоритма Apriori для Python 2.7 и 3.3-3.5, представленная как API-интерфейсы и как интерфейсы командной строки. Модуль состоит из одного файла и не зависит от других библиотек.

Код реализации алгоритма Apriori:

```
#импортируем apriori
from apyori import apriori
[
#подготавливаем данные для использования
]
#реализуем алгоритм на подготовленных данных
res = list(apriori(trscs, min_support=0.005, min_confidence=0.1,
min_lift=1, min_length=2))
```

где trscs – массив предварительно подготовленных данных, min_support – минимальное значение для параметра support, min_confidence – минимальное значение для параметра confidence, min_lift – минимальное значение для параметра lift, min_length – минимальная длина набора совместно покупаемых товаров. Более подробный код метода представлен в приложении 5.

Результаты, полученные при применении алгоритма с заданными параметрами на имеющихся данных представлены в таблице 2.

Таблица 2

Результаты анализа данных с применением алгоритма Apriori

№п/п	**Основной товар**	**Товар-компаньон**	**Support**	**Confidence**	**Lift**
1	Товар 44	Товар 29	0,0037	0,41	26,74
2	Товар 30	Товар 29	0,0031	0,22	14,36
3	Товар 3	Товар 4	0,0031	0,71	11,43
4	Товар 8	Товар 29	0,004	0,08	5,43
5	Товар 29	Товар 4	0,0034	0,26	4,16
6	Товар 17	Товар 5	0,0037	0,18	3,88

Продолжение таблицы 2. Результаты анализа данных с применением алгоритма Apriori

№п/п	Основной товар	Товар-компаньон	Support	Confidence	Lift
7	Товар 10	Товар 8	0,004	0,13	3,47
8	Товар 14	Товар 4	0,0059	0,21	3,30
9	Товар 10	Товар 4	0,0056	0,21	3,23
10	Товар 10	Товар 3	0,0056	0,19	2,76
11	Товар 8	Товар 4	0,0046	0,13	2,02

Цифра в названии товара – это порядковый номер товара согласно анализу доходности, то есть «Товар 44» находится на 44 месте по уровню приносимой прибыли.

Из таблицы видно, что было найдено 11 пар товаров, чьи значения lift больше 1. В основном, товары, входящие в пары наиболее часто покупаемых совместно, находятся в списке десяти наиболее прибыльных для отдельных товаров, однако максимальное значение lift имеет пара товаров, не входящих в топ по доходности. Параметры support и confidence имеют очень низкое значение, что в данном случае обусловлено особенностями используемого набора данных.

Для конечного пользователя приложения отображаются только столбцы «основной товар» и «товар-компаньон». На рисунке 16 представлен интерфейс пользователя, позволяющий просмотреть результаты анализа корзины.

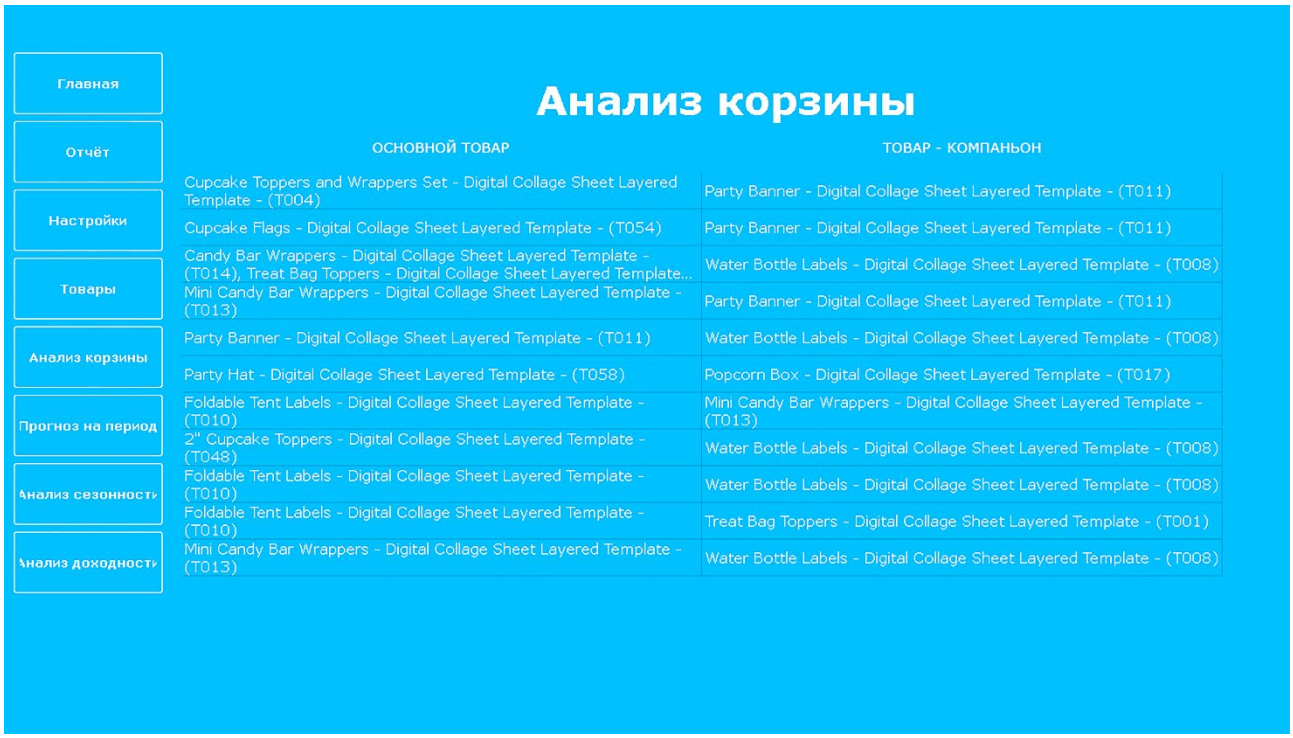

Рис. 16. Реализация интерфейса для анализа потребительской корзины

3.8. ПРОГРАММНАЯ РЕАЛИЗАЦИ ЗАДАЧИ АНАЛИЗА ДОХОДНОСТИ (ABC)

Для анализа берутся данные «Item Name», «Item Total» из отчетоа «EtsySoldOrderItems»

```
#группируем данные по «Item Name» и суммируем «Item Total»
       all_items = all_items[['Item Name', 'Item Total']]
             all_items = all_items.groupby('Item Name')['Item
     Total'].sum()
             all_items = pd.DataFrame(data=np.array([all_items.index,
all_items.values]).T, columns=['Item Name', 'Item Total'])
#суммируем и вычисляем процент доходности для каждой позиции
     summ = all_items['Item Total'].values.sum()
             all_items['total_percentage'] = all_items['Item
     Total'].values/summ
             all_items = all_items.sort_values(by='Item Total',
     ascending=False)

             all_items['cum_perc'] =
np.cumsum(all_items['total_percentage'].values)
#группируем объекты по процентам нарастающим итогом
        def group(x):
            if x < 0.8:
                return 'A'
            elif x < 0.95:
                return 'B'
            else:
                return 'C'
#создаем сводную таблицу ABC анализа
```

```
    all_items.to_csv(path + '/ABC_table_' + '_'.join([str(y) for y in
ys]) + '.csv')
```

Более подробный код метода представлен в приложении 6.

Визуально результаты представлены следующим образом:

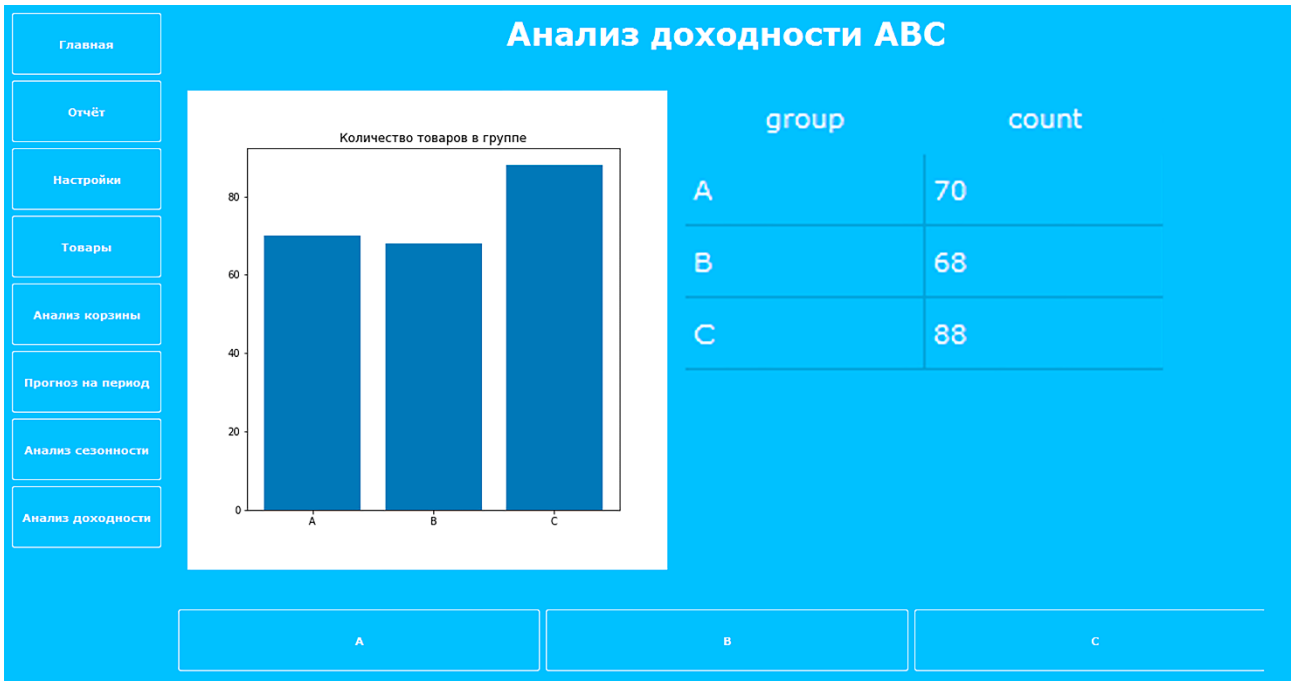

Рис. 17. Реализация интерфейса для получения результатов анализ доходности

Отдельно можно посмотреть список позиций в каждой группе с указанием названия, общей суммы и процента:

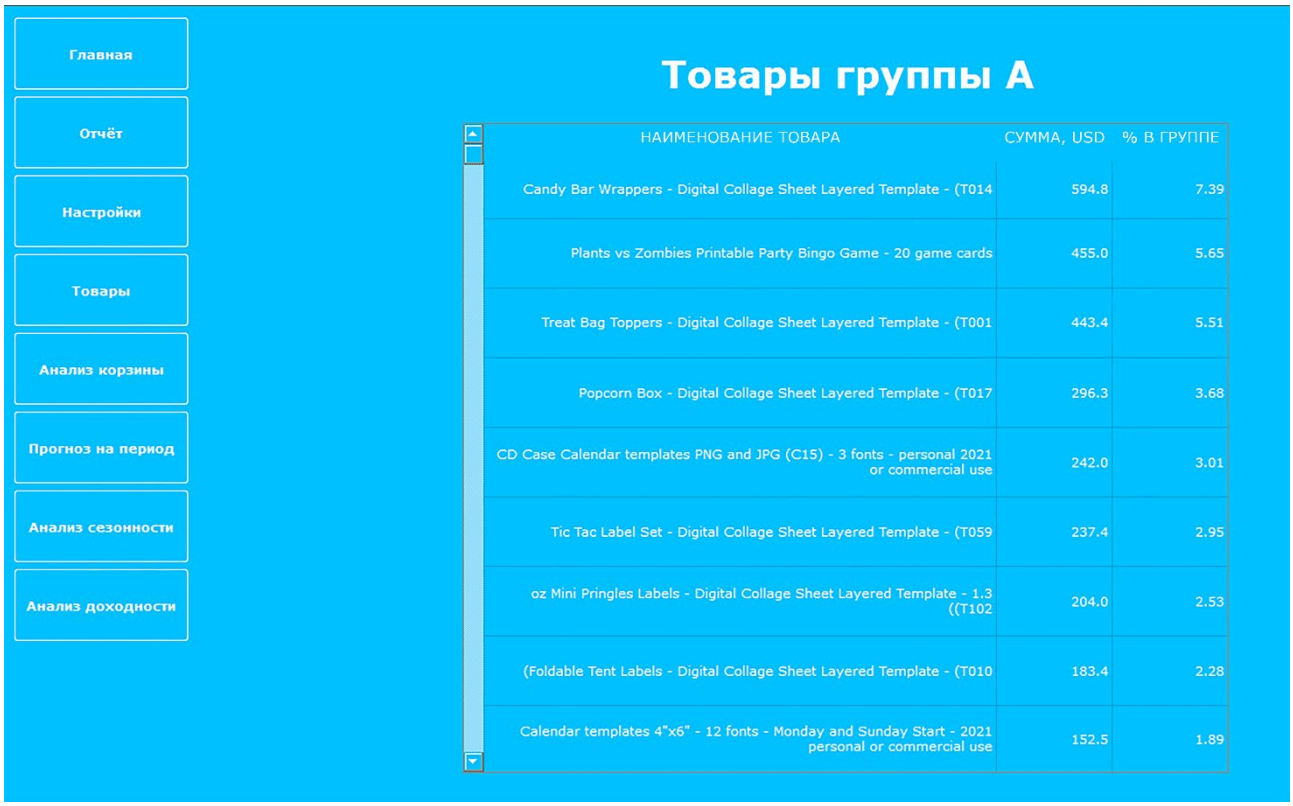

Рис. 18. Интерфейс «Товары группы А»

3.9. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ УРОВНЯ ПРОДАЖ НА ПЕРИОД 12 МЕСЯЦЕВ

Для реализации функционала прогнозирования в пользовательском интерфейсе было разработано окно «Прогноз на период». Данное окно позволяет посмотреть, какие месяцы были самыми активными в плане продаж. Показывает прогнозируемый доход за период, также прогноз по месяцам. На рисунке 19 представлен интерфейс пользователя, позволяющий просмотреть результаты прогноза на период.

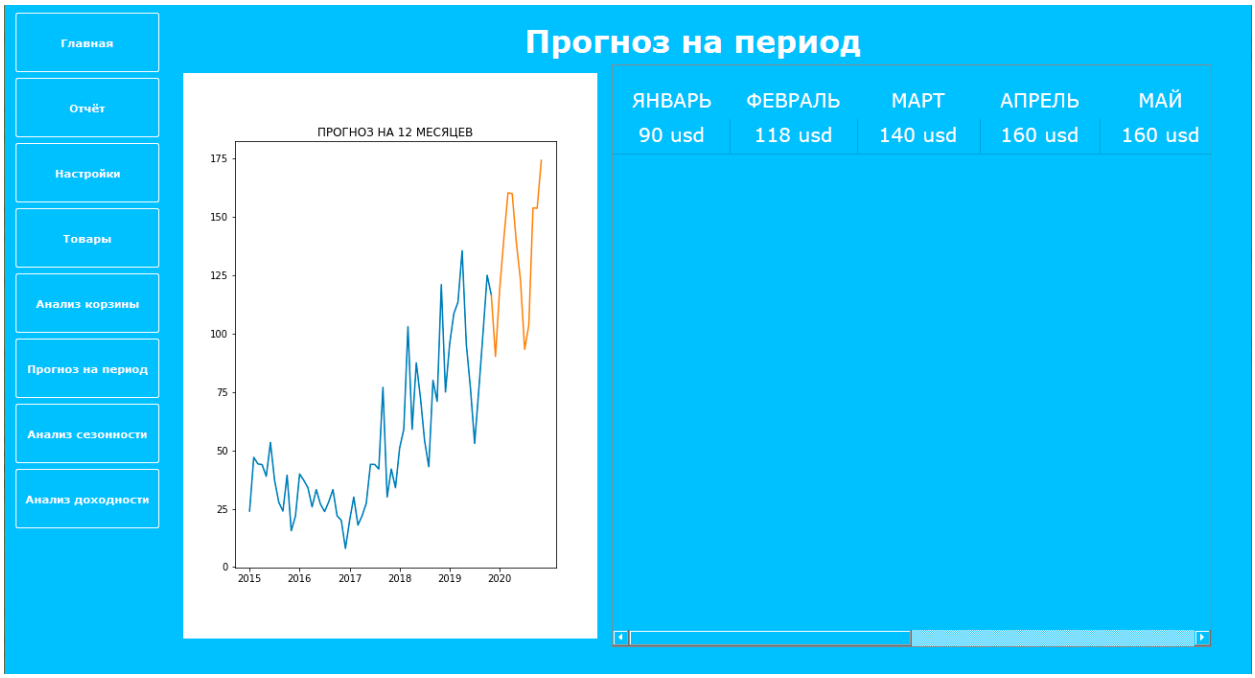

Рис. 19. Интерфейс «Прогноз на период»

Окно делится на две части: в правой таблица по месяцам с суммами прогнозируемого дохода на последующий год начиная с января в валюте магазина. В левой общий график, отображающий синим продажи за предыдущие года, желтым – прогноз на следующий календарный год. Значения по оси абсцисс – года, по оси ординат – сумма дохода.

Метод Холта-Винтерса реализует тройное экспоненциальное сглаживание, учитывающее тренд и сезонность. Есть два варианта этого метода:

- Аддитивный (additive) метод: сезонные колебания примерно постоянны на протяжении всего ряда.

- Мультипликативный (multiplicative) метод: сезонные вариации меняются пропорционально уровню ряда.

Подключаем необходимые модули:

```
import statsmodels.api as sm
import pandas as pd
```

Statsmodels – это модуль Python, который предоставляет классы и функции для оценки множества различных статистических моделей, а также для проведения статистических тестов и исследования статистических данных. Для каждого метода оценки доступен обширный список результатов статистики. Результаты проверяются на соответствие существующим статистическим

пакетам, чтобы убедиться, что они верны. Пакет выпущен под лицензией Modified BSD с открытым исходным кодом. Statsmodels поддерживает специфические модели, используя формулы языка R и массивы данных pandas.

Данные берутся из отчета EtsyListingsDownload.csv, откуда вытаскиваются все активные на данный момент листинги, и отчетов EtsySoldOrderItems за прошедшие года, откуда берутся непосредственно данные по фактическим продажам. В анализе используются только активные листинги, так как целью является получение предсказания для существующих товаров:

```
active_items = pd.read_csv(path + '/active_items_' +
       '_'.join([str(y) for y in ys]) + '.csv')
```

Данные группируются по месяцам и суммируются по стоимости:

```
active_items = active_items.groupby('month')
     ['Item Total'].sum()
```

Далее создается датафрейм и транспонируется для использования в модели:

```
active_items = pd.DataFrame(data=np.array([active_items.index,
     active_items.values]).T, columns=['month', 'Item Total'])
```

На основе созданного датафрейма формируется модель:

```
model = sm.tsa.ExponentialSmoothing(active_items
     ['Item Total'].values, seasonal_periods=12, trend='mul',
     seasonal='mul').fit()
```

Полный код метода представлен в приложении 7.

Период сезонности – 12 месяцев, тренд и сезонность рассчитываются с использованием мультипликативного метода.

На графике прогноза на 2020 год, представленном на рисунке 20, можно отследить два пика: с марта по май и с октября по декабрь.

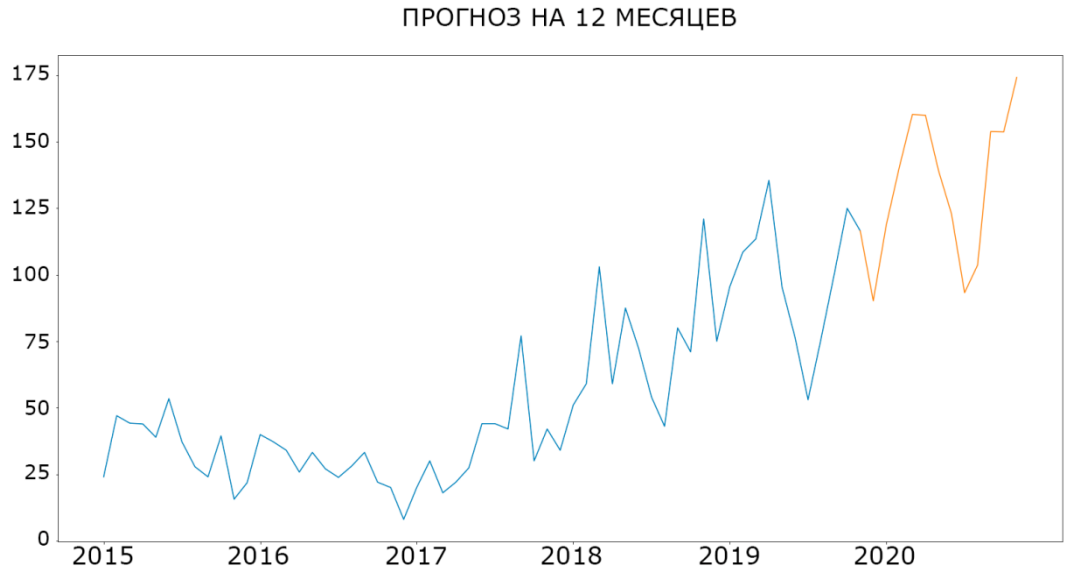

Рис. 20. График прогнозирования продаж

Для более ясного понимания прогноза эти значения приведены в таблице данных по месяцам в интерфейсе пользователя, где можно посмотреть прогнозируемую сумму дохода для каждого месяца.

При сравнении прогноза на один (2020) год и на два (2019 и 2020) года, можно увидеть, что прогнозируемый доход на 2019 год превышает показатели фактического дохода во второй половине периода. Прогноз на 2020 год имеет незначительные отличия.

Рис. 21. Сравнение прогнозных моделей на период в 12 и 24 месяца

3.10. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ЗАДАЧИ АНАЛИЗА СЕЗОННОСТИ ПРОДАЖ

Для реализации функционала анализа сезонности в пользовательском интерфейсе было разработано окно «Анализ сезонности». Данное окно позволяет отслеживать уровень продаж по годам и по сезонам. Окно предоставляет два варианта отображения уровня продаж – общий по всем товарам и отдельные по каждому активному товару.

Данные берутся из отчета EtsyListingsDownload.csv, откуда вытаскиваются все активные на данный момент листинги, и отчетов EtsySoldOrderItems за прошедшие года, откуда берутся непосредственно данные по фактическим продажам.

При помощи функции create_csv на базе отчетов EtsyListingsDownload и EtsySoldOrderItems создается общий csv файл, содержащий данные по активным товарам (см. приложение 8): название товара, год и месяц продажи, количество и сумму. Файл «active_items_(…года_продаж…)» сохраняется в общей папке с данными, указанной в настройках программы, и доступен для просмотра в любом текстовом редакторе.

Далее этот файл используется в качестве источника данных для анализа сезонности.

Общий анализ по всем активным товарам:

```
def seasonal_all (ys, path):
[…]
      active_items = active_items.groupby('month')['Item Total'].sum()
      active_items = pd.DataFrame(data=np.array([active_items.index,
active_items.values]).T, columns=['month', 'Item Total'])
   active_items = active_items.iloc[:-1]
```

Анализ по каждому отдельному товару делается по тому же принципу, также дополнительно формируется таблица с данными по месяцам:

```
def item_seasonal(ys, path, item):
[…]
      for m in range(int(active_items['month'].min()),
int(active_items['month'].max())):
            if m not in (active_items['month'].unique()):
                  active_items.loc[len(active_items)] = [m, 0.001]
      active_items = active_items.sort_values('month')
```

На рисунке 22 представлен интерфейс пользователя, где можно посмотреть отчет по всем активным товарам магазина.

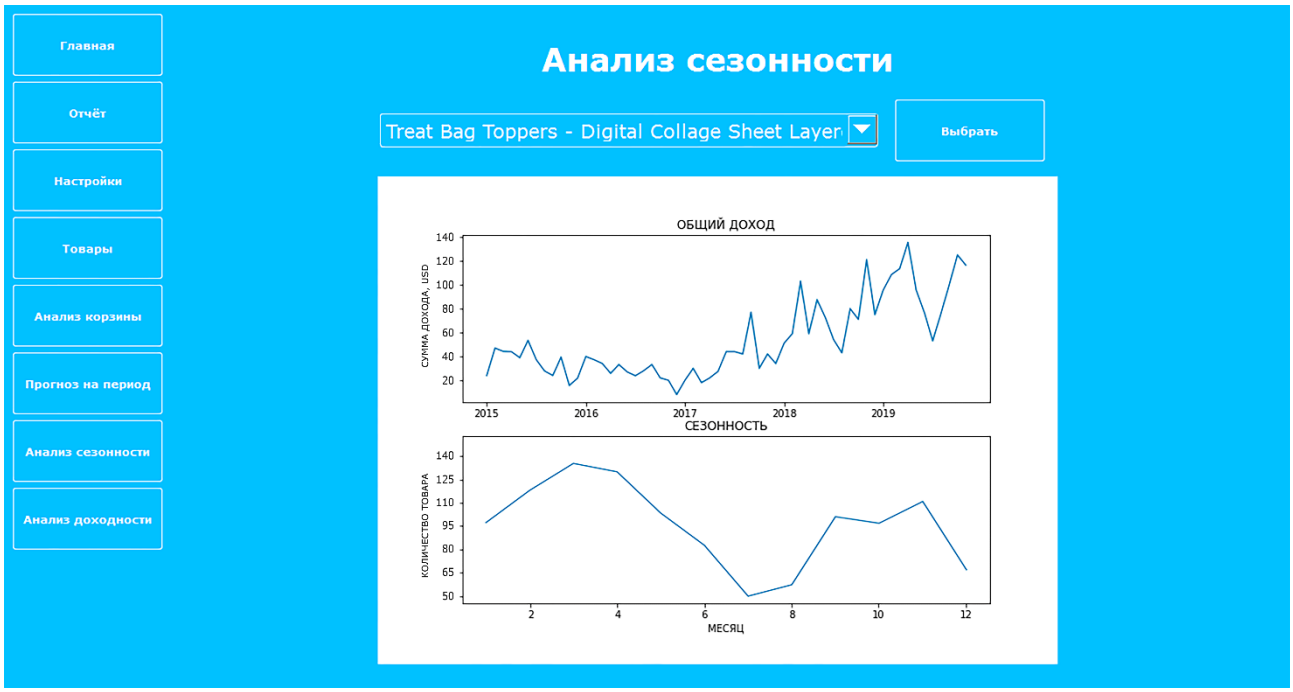

Рис. 22. Интерфейс «Анализ сезонности»

График «Общий доход» показывает доход от продаж всех активных товаров за те года, по которым имеются данные. По оси абсцисс идут года, по оси ординат – общий доход от всех активных товаров в валюте магазина. Учитываются только активные на момент анализа листинги, товары, вышедшие из продажи, не используются при анализе.

График «Сезонность» показывает средние сезонные колебания продаж активных товаров за все прошедшие года. По оси абсцисс показаны месяцы, по оси ординат – количество проданных товаров.

Второй вариант – это просмотр анализа сезонности по отдельным товарам. Выбор товара осуществляется с помощью выпадающего меню.

При выборе отдельного товара в окне отображаются два графика, аналогичных общим графикам по магазину. То есть, по каждому отдельному товару можно увидеть общий доход по годам и график средних сезонных колебаний. На рисунке 23 представлен интерфейс пользователя, предоставляющий возможность посмотреть отчет по отдельному товару.

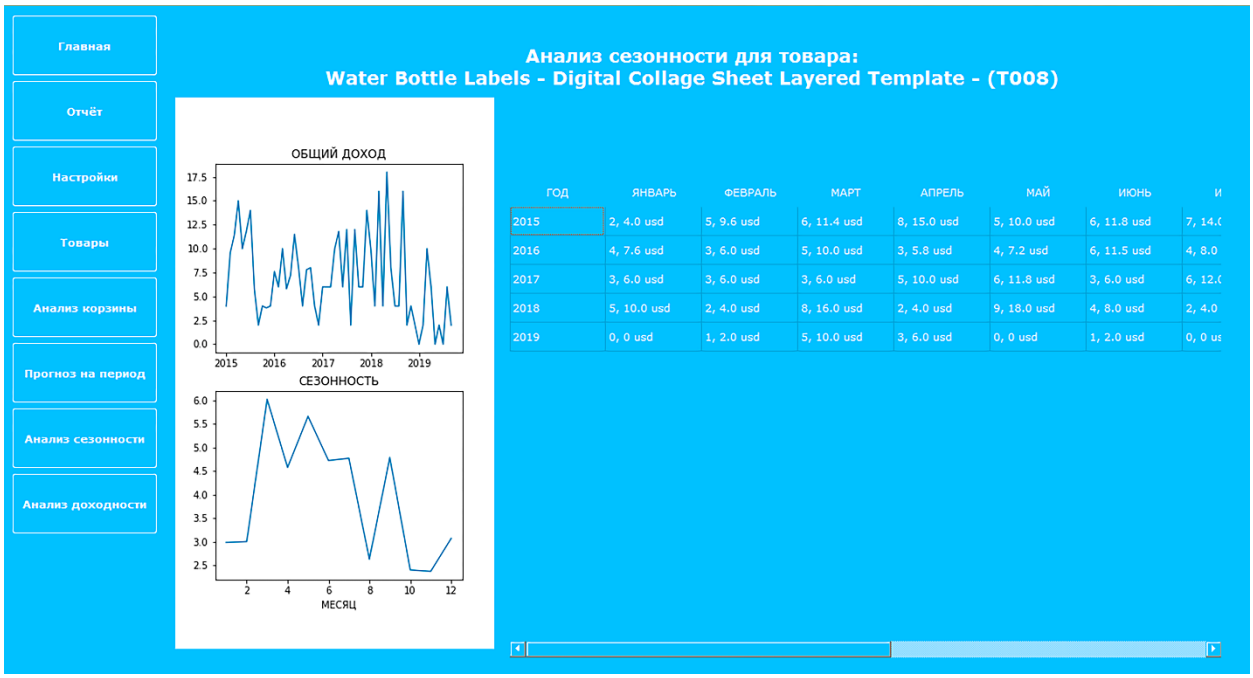

Рис. 23. Интерфейс «Анализ сезонности по отдельному товару»

В таблице на рисунке 24 приводятся данные продаж по месяцам за каждый год активности товара.

ГОД	ЯНВАРЬ	ФЕВРАЛЬ	МАРТ	АПРЕЛЬ	МАЙ
2015	2, 4.0 usd	5, 9.6 usd	6, 11.4 usd	8, 15.0 usd	5, 10.0 usd
2016	4, 7.6 usd	3, 6.0 usd	5, 10.0 usd	3, 5.8 usd	4, 7.2 usd
2017	3, 6.0 usd	3, 6.0 usd	3, 6.0 usd	5, 10.0 usd	6, 11.8 usd
2018	5, 10.0 usd	2, 4.0 usd	8, 16.0 usd	2, 4.0 usd	9, 18.0 usd
2019	0, 0 usd	1, 2.0 usd	5, 10.0 usd	3, 6.0 usd	0, 0 usd

Рис. 24. Данные продаж по годам и месяцам

Первая цифра в ячейке – это сумма проданных товаров в этом месяце, а вторая – общая сумма дохода в валюте магазина.

ЗАКЛЮЧЕНИЕ

В результате выполнения выпускной квалификационной работы было разработано приложение, ориентированное на российского пользователя, основными функциями которого являются сбор и обработка данных, предоставляемых торговой площадкой Etsy.

Были изучены основные алгоритмы поиска ассоциативных правил. Для разработки приложения был выбран алгоритм Apriori. Этот алгоритм является наиболее используемым на данный момент, так как благодаря свойству анти-монотонности дает возможность обрабатывать большие объемы данных без значительного увеличения временных затрат. Кроме того, реализация алгоритма Apriori на языке python является очень простой и компактной.

В процессе выполнения выпускной квалификационной работы были изучены различные методы прогнозирования, из которых для дальнейшего использования при разработке приложения была выбрана модель Холта-Винтерса. Модель Холта-Винтерса учитывает экспоненциальный тренд и сезонность, что очень важно при работе с данными о продажах. С учетом специфики данных интернет-магазина эта модель является хорошим выбором для кратко- и среднесрочного прогнозирования.

В результате было разработано приложение на языке python 3.8, GUI был сделан при помощи программы Qt Designer и библиотеки PyQt5. Для работы с таблицами использовались библиотеки pandas и numpy.

В качестве источника данных используются файлы отчетов магазина Etsy. Данные предоставляются в формате .csv, и предварительно должны быть скачаны в соответствующей секции аккаунта Etsy.

Разработанное приложение позволяет посмотреть основную статистическую информацию по магазину, как, например, общую сумму дохода за исследуемый период, основные товары, приносящие наибольший доход, наиболее популярные товары и так далее. Кроме того, приложение предоставляет доступ к результатам анализа потребительской корзины, что дает возможность спланировать дальнейшие действия для увеличения прибыли, например, сформировать акционные предложения. На основе результатов

анализа сезонности можно провести оптимизацию стока и спланировать работу в зависимости от активности продаж в магазине. Приложение позволяет также спрогнозировать уровень дохода на период в двенадцать месяцев.

В перспективе в приложение могут быть добавлены функции автоматического взаимодействия с сайтом Etsy и анализа текстовой части описания товара.

СПИСОК ЛИТЕРАТУРЫ

1. Исайченкова В.В., Обеспечение повышения конкурентноспособности промышленного предприятия в условиях цифровой экономики // Электронный научный журнал «Век качества». 2019. №2. С. 91-103. Научная электронная библиотека «КиберЛенинка» .NET: [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/obespechenie-povysheniya-konkurentosposobnosti-promyshlennogo-predpriyatiya-v-usloviyah-tsifrovoy-ekonomiki/viewer (дата обращения: 14.10.20)
2. Трачук А.В., Линдер Н.В. Распространение инструментов электронного бизнеса в России: результаты эмпирического исследования // Российский журнал менеджмента. 2017. Том 15. №1. С. 27-50. Научная электронная библиотека eLibrary .NET: [Электронный ресурс]. URL: https://www.elibrary.ru/download/elibrary_29245891_64714333.pdf (дата обращения: 22.10.2020)
3. Лисовский А.Л. Оптимизация бизнес-процессов для перехода к устойчивому развитию в условиях четвёртой промышленной революции // Стратегические решения & риск-менеджмент. 2018. №4. С.10-17. Научная электронная библиотека «КиберЛенинка» .NET: [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/optimizatsiya-biznes-protsessov-dlya-perehoda-k-ustoychivomu-razvitiyu-v-usloviyah-chetvertoy-promyshlennoy-revolyutsii/viewer (дата обращения: 22.10.2020)
4. Дмитриенко И., Рукодельники XXI века // Еженедельный журнал Профиль .NET: [Электронный ресурс]. URL: https://profile.ru/society/rukodelniki-xxi-veka-3331/ (дата обращения: 15.10.20)
5. О бизнесе в стиле hande made // Новоделие: социокультурные проекты .NET: [Электронный ресурс]. URL: https://www.novodelye.ru/blog/profil_o_hand_made_biznese/ (дата обращения: 02.06.2020)
6. Etsy - Statistics & Facts // Statista official site .NET: [Электронный ресурс]. URL: https://www.statista.com/topics/2501/etsy/ (дата обращения: 02.06.2020).

7. Чернышова Г. Ю., Самаркина Е. А. Методы интеллектуального анализа данных для прогнозирования финансовых временных рядов // Изв. Сарат. ун-та. Нов. сер. Сер. Экономика. Управление. Право. 2019. Т. 19, вып. 2. С. 181–188.
8. Чубукова, И. А. Data Mining : учебное пособие // Москва : Интернет-Университет Информационных Технологий (ИНТУИТ) : Бином. Лаборатория знаний, 2008. 2-е изд., испр. 383 с.
9. Дюк В.А. Data Mining – интеллектуальный анализ данных // OLAP и Business Intelligence .NET: [Электронный ресурс]. URL: http://www.olap.ru/basic/dm2.asp. (дата обращения: 22.10.2020)
10. eRank // eRank official website .NET: [Электронный ресурс]. URL: https://erank.com/ (дата обращения: 14.04.2021)
11. Vela // Vela official website .NET: [Электронный ресурс]. URL: https://welcome.getvela.com/ (дата обращения: 14.04.2021)
12. Craftybase // Craftybase official website .NET: [Электронный ресурс]. URL: https://craftybase.com/ (дата обращения: 14.04.2021)
13. Marmalead // Marmalead official website .NET: [Электронный ресурс]. URL: https://marmalead.com/ (дата обращения: 14.04.2021)
14. Putler // Putler official website .NET: [Электронный ресурс]. URL: https://www.putler.com/integrations/etsy/ (дата обращения: 14.04.2021)
15. Usman Malik. Association Rule Mining via Apriori Algorithm in Python // Stack Abuse .NET: [Электронный ресурс]. URL: https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python (дата обращения: 17.04.2021)
16. FP Growth: Frequent Pattern Generation in Data Mining with Python Implementation // Towards Data Science .NET: [Электронный ресурс]. URL: https://towardsdatascience.com/fp-growth-frequent-pattern-generation-in-data-mining-with-python-implementation-244e561ab1c3 (дата обращения: 17.04.2021)
17. ML ECLAT // GeeksForGeeks .NET: [Электронный ресурс]. URL: Algorithm https://www.geeksforgeeks.org/ml-eclat-algorithm/ (дата обращения: 17.04.2021)

18. Apyori 1.1.2 // The Python Package Index .NET: [Электронный ресурс]. URL: https://pypi.org/project/apyori/(дата обращения: 17.04.2021)
19. ABC анализ. Что это и как можно использовать? // Marketing Education .NET: [Электронный ресурс]. URL: https://maed.ru/podvodnye-kamni-abc-analiza-kak-izbezhat-oshibok-pri-ego-provedenii/ (дата обращения: 21.04.2021)
20. Правило Парето (закон Парето) // Записки маркетолога .NET: [Электронный ресурс]. URL: http://www.marketch.ru/marketing_dictionary/marketing_terms_p/pareto_rule_law/ (дата обращения: 21.04.2021)
21. Корецкая Т.В., Краткосрочное прогнозирование комплексных переменных с использованием метода Брауна. Вестник ОГУ. №11. Ноябрь, 2008. С.121-126
22. Дядичев В.В., Ромашка Е.В., Голуб Т.В., Геополитика и экогеодинамика регионов. Том 1 (11). Вып. 3. 2015. С. 23–29
23. Модель скользящего среднего // Bstudy .NET: [Электронный ресурс]. URL: https://bstudy.net/708118/ekonomika/model_skolzyaschego_srednego (дата обращения: 24.04.2021)
24. Носко В.П., Эконометрика. Введение в регрессионный анализ временных рядов // Москва. – 2009. 273 с.
25. Метод наименьших квадратов // Matematicus .NET: [Электронный ресурс]. URL: https://www.matematicus.ru/matematicheskaya-statistika/metod-naimenshih-kvadratov-regressiya (дата обращения: 24.04.2021)
26. Модель Брауна // Studbooks .NET: [Электронный ресурс]. URL: https://studbooks.net/2397606/matematika_himiya_fizika/model_brauna (дата обращения: 24.04.2021)
27. Авторегрессионные модели (AR) // Univer-nn .NET: [Электронный ресурс]. URL: https://univer-nn.ru/avtoregressionnye-modeli-arp/ (дата обращения: 24.04.2021)
28. Модель авторегрессии скользящего среднего // Энциклопедия по экономике .NET: [Электронный ресурс]. URL: https://economy-ru.info/info/75662/ (дата обращения: 24.04.2021)

29. The Box-Jenkins Method // NCSS: Statistical, Graphics, and Sample Size Software .NET: [Электронный ресурс]. URL: https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/The_Box-Jenkins_Method.pdf (дата обращения: 24.04.2021)
30. Sachin Date., Holt-Winters Exponential Smoothing 2020 // Towards Data Science .NET: [Электронный ресурс]. URL: https://towardsdatascience.com/holt-winters-exponential-smoothing-d703072c0572 (дата обращения: 09.02.2021)
31. What is PyQt? // Riverbank Computing .NET: [Электронный ресурс]. URL: https://riverbankcomputing.com/software/pyqt (дата обращения: 17.08.2020)