

Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

УДК 339.37:519.6

На правах рукописи

ЛИТВИН
Вероника Валерьевна

**МАТЕМАТИЧЕСКИЕ МЕТОДЫ И ИНСТРУМЕНТАЛЬНЫЕ
СРЕДСТВА АНАЛИЗА ДАННЫХ ПРЕДПРИЯТИЯ РОЗНИЧНОЙ
ТОРГОВЛИ**

АВТОРЕФЕРАТ
магистерской диссертации на соискание степени
магистра экономических наук

по специальности 1-25 80 08 «Математические и инструментальные
методы экономики»

Минск 2018

Работа выполнена на кафедре экономической информатики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Научный руководитель: **Алёхина Алина Энодиевна,**
кандидат экономических наук, доцент кафедры экономической информатики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Рецензент: **Синявская Ольга Александровна,**
кандидат экономических наук, доцент кафедры промышленного маркетинга и коммуникаций учреждения образования «Белорусский государственный экономический университет»

Защита диссертации состоится «30» января 2018 г. года в 10⁰⁰ часов на заседании Государственной комиссии по защите магистерских диссертаций в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» по адресу: 220013, г.Минск, ул. Платонова, 39, 5 уч. корп., ауд. 806, тел.: 293-89-92, e-mail: kafei@bsuir.by.

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

СОГЛАСОВАНО

Научный руководитель
канд.экон.наук, доцент

А.Э. Алёхина

ВВЕДЕНИЕ

Повсеместное использование компьютеров привело к пониманию важности задач, связанных с анализом накопленной информации для извлечения своевременных новых знаний. Обострение рыночной конкуренции предъявляют новые требования к информационно-аналитическому обеспечению системы управления. Возникла потребность в создании хранилищ данных и систем поддержки принятия решений, основанных, в том числе, и на методах теории искусственного интеллекта.

Для эффективного управления на предприятиях возникла необходимость принятия решений, подсказывающих изменения, а не запаздывающие реакции. Управление предприятием, банком, различными сферами бизнеса, немыслимо без процессов анализа, выявления определенных закономерностей и зависимостей, прогнозирования тенденций и рисков.

При увеличении и усложнении процессов анализа возникла проблема интерпретации данных анализа. Результаты, получаемые при анализе данных не всегда удобны для восприятия человеком, так как достаточно сложно быстро и легко найти новые и полезные знания. В связи с этим возникает необходимость в дополнительных средствах отображения результатов анализа.

Основной идеей средств визуального анализа данных является представление данных в некоторой визуальной форме, позволяющей человеку погрузиться в данные, работать с их визуальным представлением, понять их суть, сделать выводы и напрямую взаимодействовать с данными.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Актуальность темы исследования магистерской диссертации обусловлена с растущим интересом в мировом IT-сообществе к анализу и визуализации больших объемов данных.

В современном мире анализ данных является важным аспектом любой компании. Накопленный объем данных неуклонно растет на предприятиях и в организациях, и объем данных увеличивается экспоненциально. Согласно новому докладу IBM Marketing Cloud, «10 ключевых тенденций в маркетинге на 2017 год», 90% данных в мире сегодня создано за последние два года. Большой объем информации делает поиск решений сложной задачей. Таким образом для достижения максимальной эффективности компании, несмотря на все изменения, должен проводиться надлежащий анализ в нужное время.

Актуальный анализ данных имеет важное значение для поддержания конкурентоспособности компании. Он дает возможность сформулировать способы улучшения бизнес-стратегий, бизнес-операций и принятия разумных бизнес-решений для достижения итогового результата компании. Крайне

важно понять ключевые области, чтобы помочь бизнесу увеличить доходы и сократить ненужные расходы.

Цель и задачи исследования

Целью данной диссертационной работы является повышение эффективности деятельности предприятия розничной торговли за счет создания визуально-аналитического приложения, использующего методы интеллектуального анализа, для принятия управленческих решений

Для осуществления цели необходимо решить следующие **задачи**:

1. Исследовать математические методы анализа данных, в частности методы интеллектуального анализа данных предприятия розничной торговли; рассмотреть возможные средства для визуализации аналитических данных.
2. Разработать эконометрические модели основных показателей провести кластерный анализ данных предприятия розничной торговли.
3. Разработать визуальное-аналитическое приложение объединяющее математических средства аналитики и оптимальное графического представления результатов анализа данных предприятия розничной торговли.

В диссертации используются такие **методы исследования**, как анализ, классификация, сравнение, кластеризация, моделирование.

Область исследования. Содержание диссертационной работы соответствует образовательному стандарту высшего образования второй ступени (магистратуры) специальности 1-25 80 08 «Математические и инструментальные методы экономики».

Теоретическая и методологическая основа исследования

В основу диссертации легли результаты известных исследований российских и зарубежных исследователей в области анализа и визуализации данных.

Для проведения анализа данных предприятия были использованы методы статистического анализа, так же метод k-средних кластерного анализа.

На одном из этапов моделирования информационной системы использовалась CASE-технология UML. Для разработки визуально-аналитического приложения использовался инструмент анализа и визуализации Microstrategy.

Информационная база исследования для интеллектуального анализа сформирована на основе статистических данных предприятия розничной торговли.

Основные положения, выносимые на защиту

1. Обзор математических методов интеллектуального анализа данных, анализ средств визуализации данных для предприятия розничной торговли.

2. Эконометрические модели основных показателей деятельности предприятия и кластерный анализ данных предприятия.

3. Визуально-аналитического приложение, предоставляющее аналитические данные в интуитивно понятном и интерактивном виде.

Теоретическая значимость диссертационной работы заключается в разработке подхода к анализу данных предприятия розничной торговли с использованием средств визуализации анализа данных.

Практическая значимость диссертационной работы заключается в разработке визуально-аналитического приложения для отображения результатов анализа данных предприятия розничной торговли.

Апробация и внедрение результатов исследования

Результаты исследования представлены на Международной научно-практической конференции «Актуальные вопросы взаимодействия образования, науки и бизнеса» (Инновационный научный центр «Импульс», Москва, 2018) и международной научно-практической конференции «Экономика и менеджмент: теоретико-методологические подходы и практические результаты исследований» (НИЦ «Поволжская научная корпорация», Самара, 2018).

Публикации

Основные положения работы и результаты диссертации изложены в двух опубликованных работах и двух в печати общим объемом 0,5 п.л. (авторский объем 0,5 п.л.).

Структура и объем работы. Структура диссертационной работы обусловлена целью, задачами и логикой исследования. Работа состоит из введения, трёх глав и заключения, библиографического списка и приложений. Общий объем диссертации – 82 страницы. Работа содержит 4 таблицы, 45 иллюстраций, 2 приложения. Библиографический список включает 51 наименование.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** рассмотрено современное состояние проблемы анализа, интерпретации и визуализации данных, указаны основные направления исследований, проводимых по данной тематике, а также описано обоснование актуальности темы диссертационной работы.

В **общей характеристике работы** показана актуальность проводимых исследований, степень разработанности проблемы, сформулированы ее цель

и задачи, представлены положения, выносимые на защиту, приведены сведения о личном вкладе соискателя, апробации результатов диссертации и их опубликованность, а также, структура и объем диссертации.

В первой главе рассматриваются проблемы, возникающие при подготовке и анализе данных предприятия. Рассматривается особый вид анализа – Data Mining или интеллектуальный анализ. Отмечены суть и цель технологии интеллектуального анализа данных.

Так же отмечено, что Data Mining развивается на стыке таких дисциплин, как математика, статистика, теория информации, машинное обучение, теория баз данных, программирование, параллельные вычисления. Решения, предлагаемые интеллектуальным анализом можно использовать для достижения необходимых результатов в развитии предприятия и успешного продвижения товаров на рынке.

Интеллектуальный анализ данных может проводиться с помощью программных продуктов следующих классов:

- специализированных программных продуктов для интеллектуального анализа;
- математических пакетов;
- электронных таблиц (и различного рода надстроек над ними);
- средств, интегрированных в системы управления базами данных;
- других программных продуктов.

Данные, которые накапливают предприятия и организации в базах данных и прочих источниках, имеют свои особенности. Бизнес данные редко накапливаются для решения задач анализа. Предприятия и организации собирают данные для коммерческих целей: ведение учета, проведения финансового анализа, составления отчетности, принятия решений и т.д. Таким образом бизнес-данные отличаются от экспериментальных данных, которые собираются для исследовательских целей. Т.е. бизнес-данные, как правило, содержат ошибки, аномалии, противоречия и пропуски.

В связи с этим рассмотрена новая методика Knowledge Discovery in Databases (KDD). Она описывает последовательность действий, которые необходимо выполнить. KDD включает в себя вопросы: подготовки данных, выбора информативных признаков, очистки данных, применения методов интеллектуального анализа, постобработки данных и интерпретации полученных результатов.

Анализ Data Mining показал, что существуют несколько основных типов задач интеллектуального анализа: классификации, задачи в которых установление зависимости дискретной выходной переменной от входных параметров; регрессии, задачи в которых установление зависимости непрерывной выходной переменной от входных переменных; кластеризации, задачи в которых заключается в делении множества объектов на группы (кластеры) схожих по параметрам и задачи ассоциаций в которых, выявление закономерностей ждущими связанными событиями.

Рассмотрены модели интеллектуального анализа и произведена классификация задач в зависимости от типа моделей. К предсказательным моделям отнесены модели классификации и последовательностей. К описательным моделям отнесены регрессионные модели, модели кластеров, модели исключений, итоговые модели, ассоциативные модели.

Более подробно в первой главе рассмотрены задачи классификации, а именно ABC-XYZ-анализ. Рассмотрены алгоритмы ABC и XYZ анализов и алгоритм совмещенного ABC-XYZ анализа. Сделаны выводы и рекомендации по каждой группе товаров. Отмечено что ABC-XYZ-анализ позволяет в полной мере оценить весь ассортимент компании и понять, как можно управлять теми или иными группами и категориями товаров, какие товары необходимо вывести из ассортимента, а для каких товаров наоборот необходимо увеличить страховой запас. Так же что ABC-XYZ-анализ позволяет улучшить качество управления запасами, что в свою очередь обеспечивает более эффективную деятельность компании в целом.

Рассмотрены более подробно задачи кластеризации. Отмечено, что достоинством кластерного анализа в том, что он позволяет осуществлять разбиение объектов не по одному параметру, а по целому набору признаков. Кроме того, кластерный анализ, в отличие от большинства математико-статистических методов, не накладывает никаких ограничений на вид рассматриваемых объектов и позволяет рассматривать множество исходных данных практически произвольной природы. Это имеет большое значение, например, для прогнозирования конъюнктуры, при наличии разнородных показателей, затрудняющих применение традиционных эконометрических подходов.

Среди задач кластеризации выделены: изучение данных, облачение анализа, сжатие данных, прогнозирование, обнаружение аномалий. Было отмечено, что в розничной торговле задачи кластеризации чаще всего применяются для облегчения анализа. Так же был приведен пример графического отображения результатов кластерного анализа.

При анализе средств визуализации для аналитических данных были описаны особенности систем Business Intelligence так же было выявлено большое разнообразие на современном рынке визуальных программных обеспечений, которые отличаются между собой инфраструктурой, управлением данными, анализом и созданием контекста, совместным поиском, поддержкой системы, стоимостью и др. показателями.

Для классификации различных платформ на рынке использовались результаты «магического квадрант» Gartner (рисунок 1).

«Магический Квадрант» распределяет компании или вендоры по секторам: Лидеры, Провидцы, Претенденты, Нишевые игроки.

Сделаны общие выводы по главе.



Рисунок 1 – «Магический квадрант» Gartner за 2016

Во второй главе приведен общий анализ предприятия розничной торговли и положительные тенденции развития предприятия. Показана необходимость более подробного анализа данных для поддержания конкурентоспособности предприятия на рынке.

Выполнен кластерный анализ методом k-средних. Анализ выполнялся над данными описывающими сферу деятельности компании, приносящую наибольшую прибыль т.е. продажу электроники.

На основании анализа были выделены 4 группы:

- Sharp и Sony;
- Hewlett Packard и Panasonic;
- Harman Kardon и RCA;
- 3Com, GPX, Hitachi, Microsoft, Minolta, New Media, Olympus, Pioneer, PNY, Xerox.

Результат разбиения данных на кластеры можно увидеть на рисунке 2.

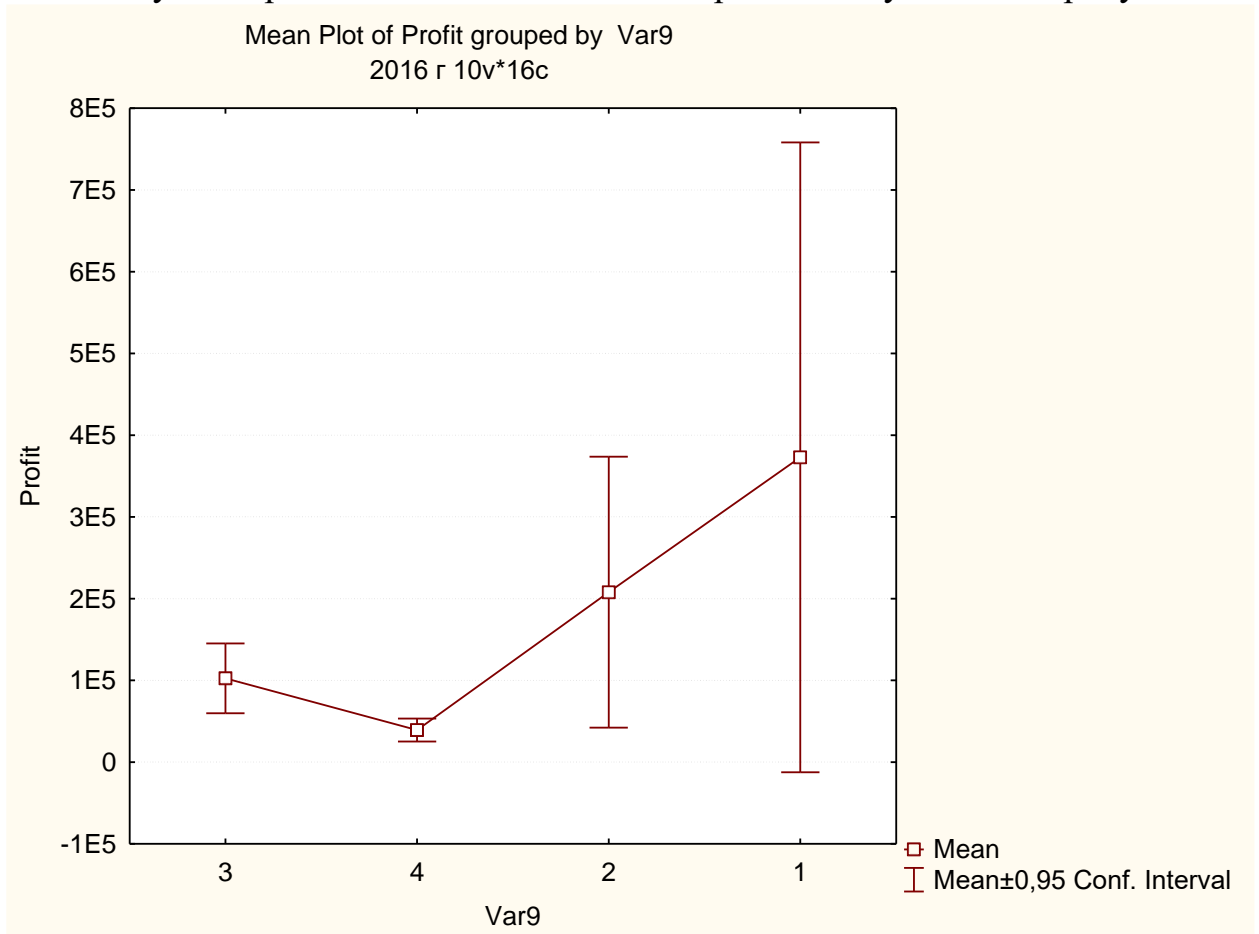


Рисунок 2 – График средних значений каждого кластера

После разбиения данных на кластеры был проведен дополнительный сравнительный анализ. Для его проведения была сформирована таблица, содержащая информацию о принадлежности бренда кластерам, прибыль по бренду и прибыль по бренду за единицу продукции.

Было отмечено, что самым прибыльным брендом является бренд первого кластера – Sony. Так же первому кластеру принадлежит бренд Sharp, что обусловлено популярностью этого бренда в США. Бренды RCA и Harman Kardon за 2016 год принесли почти одинаковую величину прибыли, однако прибыль за единицу продукции между этими брендами отличается в почти в 5 раз. На основании предоставленных данных сделан вывод, что компании необходимо обратить особое внимание на бренд Harman Kardon. Потому что бренд и не пользуется особой популярностью среди клиентов. Принадлежность его 3 кластеру обусловлена только высокой стоимостью продукции за единицу товара. Бренд нуждается в постоянном мониторинге результатов эффективности и выводе из ассортимента при необходимости. Наименее прибыльным и популярным брендом за 2016 год, оказался бренд Microsoft.

Были построены и проанализированы эконометрические модели брендов.

Была рассмотрена динамика показателя брендов на графике (рисунок 3). На основании данного графика динамики показателя (рассматривается динамика прибыли бренда 3Com) предположено, что модель представляет собой декомпозицию детерминированных составляющих: тренда и сезонности, а также случайной компоненты. Об этом свидетельствуют систематические колебания, описывающие сезонные изменения, относительно возрастающей линии тренда.

Постоянный размах колебаний относительно тенденции свидетельствует в пользу аддитивной модели.

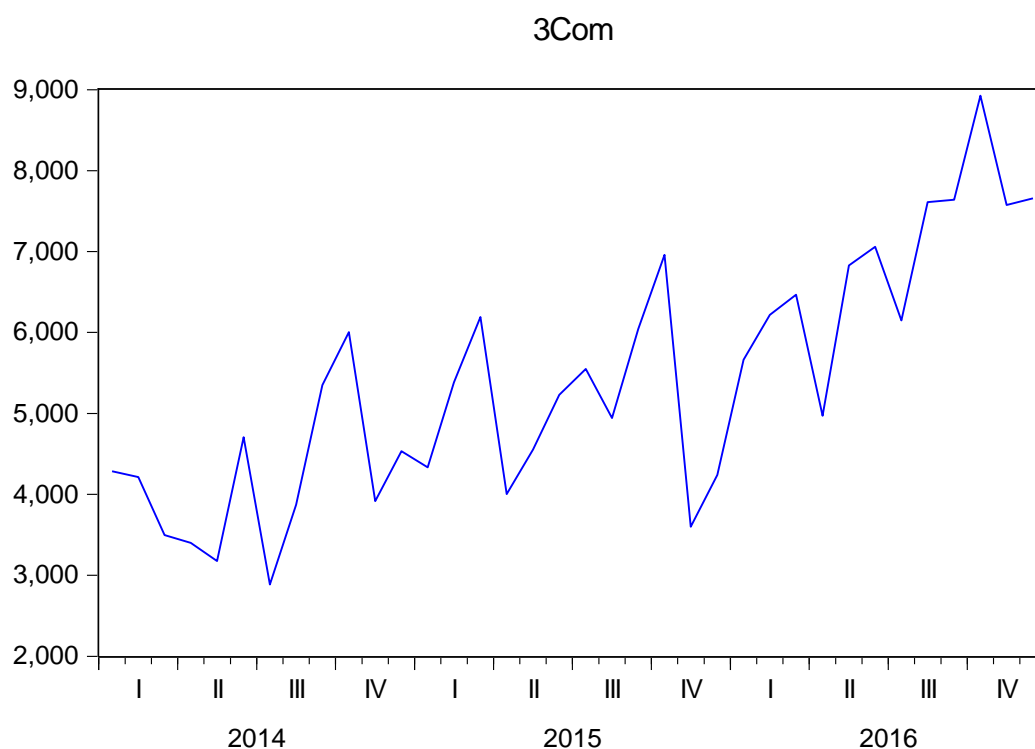


Рисунок 3 – Динамика показателя прибыли для бренда 3Com

Была определена спецификация уравнения эконометрической модели (1):

$$X_t = T(t) + S(t) + eps, \quad (1)$$

где $T(t)$ – линейный тренд,

$S(t)$ – сезонная компонента, моделируемая фиктивными переменными,

eps – случайная ошибка.

Для анализируемых данных была получена следующая спецификация уравнения (2):

$$pr_1 = 3879 + 63,83t - 1008,13s(4)_t + 1485,52s(10)_t - 909s(11)_t + 98,39dt(8,2015) + \varepsilon_t, \quad (2)$$

где t – тренд,

$s(4)$ – фиктивная переменная, равная 1 для апреля и 0 во всех остальных случаях,

$s(10)$ – фиктивная переменная, равная 1 для октября,

$s(11)$ – фиктивная переменная, равная 1 ноября.

Отсутствие сезонных компонент для ряда периодов свидетельствует о недостоверном статистическом различии в данных месяцах по отношению к эталонной категории, определенной в модели как декабрь.

Представленные коэффициенты являются статистически значимыми на 5%-уровне, о чем говорят высокие значения соответствующих t-статистик. В целом модель является высоко значимой (F статистика равна 26,42 p-value < 0,05). Коэффициент детерминации – 0,81, скорректированный коэффициент детерминации – 0,78.

Модель удовлетворяет всем модельным предположениям. Анализ остатков на нормальное распределение проведен с помощью гистограммы и критерия Жака-Бера. Остатки подчиняются нормальному закону распределения: JB = 1,55, p > 0,4 (рисунок 4).

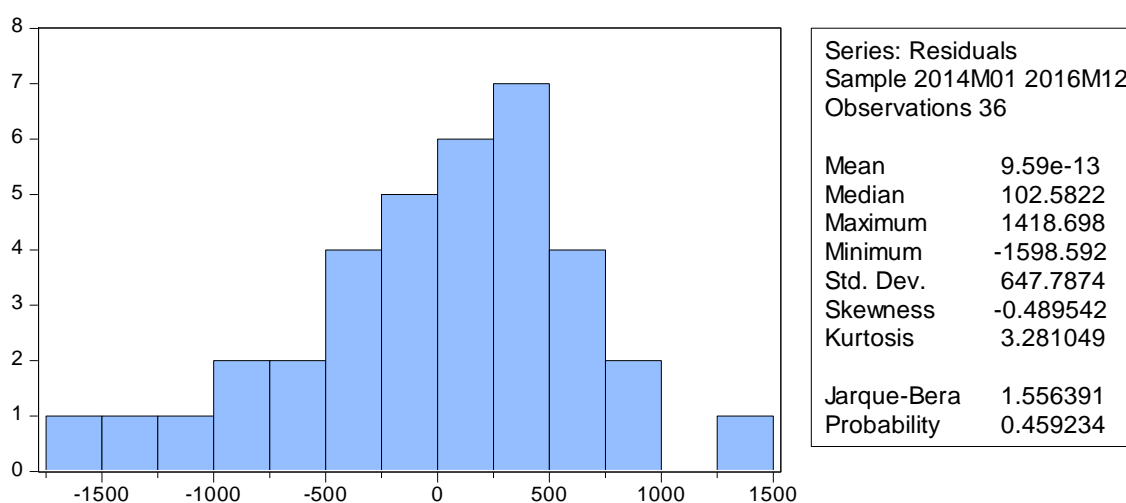


Рисунок 4 – Анализ остатков бренда 3Com на нормальное распределение

Отсутствует гетероскедастичность в остатках, о чем свидетельствует F-статистика теста Уайта, равная 0,48, p > 0,91. Остатки модели не автокоррелированы на основании теста Бреуша-Годфри: F = 0,13, p > 0,87.

Таким образом сделан общий вывод о том, что по совокупности статистических характеристик модель может быть признана удовлетворительной и может быть использована в дальнейшем для прогнозирования показателей в будущем

Сделаны общие выводы по главе.

В третьей главе представлены результаты разработки визуально-аналитического приложения. В процессе работы над диссертацией было разработано визуально-аналитическое приложение при помощи средств Business Intelligence. Данное визуально-аналитическое приложение, или дашборд, является синтезом мощных математических средств аналитики и оптимального графического представления результатов анализа. Дашборд предоставляет информацию для руководства в понятном компактном виде, а также позволяет интерактивно изменять различные параметры анализа данных.

Было отмечено, что получаемая информация будет впоследствии использоваться руководством для принятия управленческих решений.

Произведено функциональное моделирование, которое более подробно рассматривает процесс анализа информации о деятельности компании. Функциональный блок разбивается на несколько процессов: работа с фильтрацией, работа с аналитической информацией, осуществление выгрузки данных в excel, внесение важной информации.

Входом функционального блока является необходимость принятия управленческого решения и текущая информация о компании. Реализация процесса (механизм) происходит при помощи аналитика компании и разработанного визуально-аналитического приложения, которое он использует. В результате процесса использования приложения получаем необходимую информацию для принятия управленческого решения и так же excel файлы с необходимой выгруженной информацией.

Произведено информационное моделирование базы данных. Это позволило разработать информационную систему в короткие сроки и с уменьшенными затратами, сделать её легко сопровождаемой и управляемой. В проекте было реализовано 8 сущностей

При помощи языка UML была разработана диаграмма вариантов использования, которая является исходным концептуальным представлением или концептуальной моделью системы в процессе ее проектирования и разработки.

Представлена разработанная система и описаны ее возможности. На рисунке 5 представлена одна из аналитических панелей разработанного визуально-аналитического приложения для анализа данных предприятия розничной торговли.

Разработанное визуально-аналитическое-приложение позволяет:

- предоставлять всю аналитическую информацию по деятельности предприятия розничной торговли в виде графиков, гистограмм, интерактивных виджетов, карт.
- осуществлять интерактивные изменения показателей для сегментации показателей;
- осуществлять дриллинг данных;
- осуществлять сортировки и фильтрации данных по брендам, по временным интервалам;

- осуществлять выгрузку данных в excel-документы;
- вносить корректирующую информацию, на которую стоит обратить внимание сотрудникам;
- предоставлять информацию через браузер и мобильное приложение.

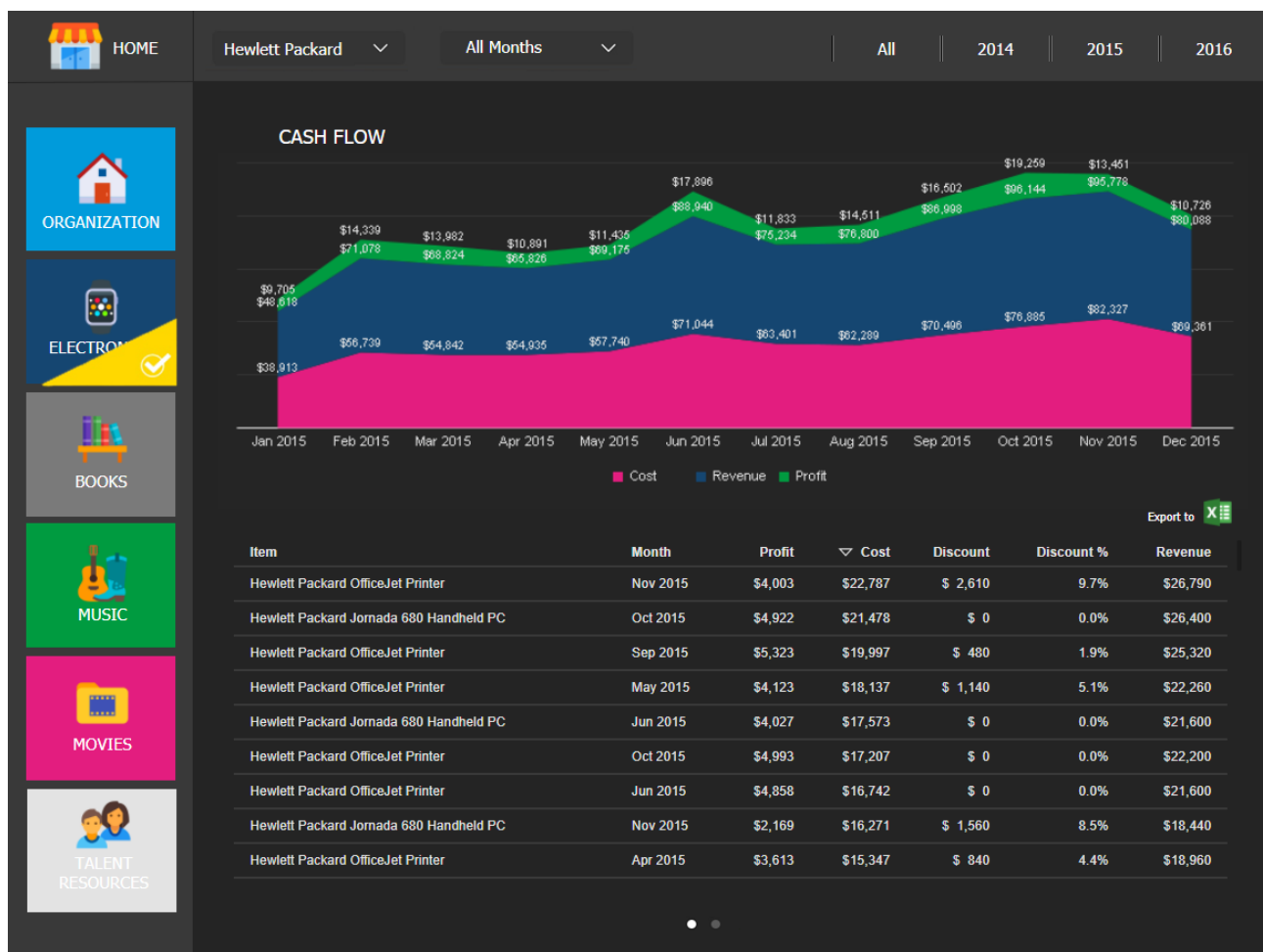


Рисунок 5 – Анализ движения денежных потоков

Сделаны общие выводы по главе.

ЗАКЛЮЧЕНИЕ

В соответствии с поставленными задачами данной диссертационной работы, можно сделать следующие выводы:

1. Проведена систематизация основных методов и задач интеллектуального анализа данных предприятия розничной торговли. Рассмотрены способы и средства визуализации аналитических данных предприятия. Структурированы данные по всем средствам визуализации на современном рынке.

2. Проведен эконометрический и кластерный анализ данных предприятия розничной торговли, проанализированы показатели и отмечены места, требующие пристального внимания компании.

3. Разработано визуально-аналитическое приложение, которое позволяет проанализировать основные аналитические показатели предприятия и предпринять соответствующие управленческие решения, на основании этих данных. Интерфейс приложения является интерактивным и интуитивно понятен любому пользователю.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

1-А. Литвин, В.В. ABC-XYZ-анализ торгового предприятия / В.В. Литвин // Экономика и менеджмент: теоретико-методологические подходы и практические результаты исследований: материалы Междунар. науч.-практич. конф / НИЦ «Поволжская научная корпорация». – Самара, 2018. – С. 56. с.

2-А. Литвин, В.В. Кластеризация как метод интеллектуального анализа: основные характеристики и визуализация / В.В. Литвин // Развитие социального и научно-технического потенциала общества: материалы Междунар. науч.-практич. конф., / Инновационный научный центр Импульс. – Москва, 2018. – С. 128. с.

РЕЗЮМЕ

Литвин Вероника Валерьевна

Математические методы и инструментальные средства анализа данных предприятия розничной торговли

Ключевые слова: интеллектуальный анализ, модели интеллектуального анализа, задачи интеллектуального анализа, извлечение знаний из баз данных, визуализация данных, средства визуализации.

Цель работы: исследование математических методов анализа данных для предприятия розничной торговли, а также использования инструментальных средств для визуализации предоставленных аналитических данных.

Полученные результаты и их новизна: проведена систематизация основных методов и задач интеллектуального анализа данных предприятия розничной торговли. Рассмотрены способы и средства визуализации аналитических данных предприятия. Структурированы данные по всем средствам визуализации на современном рынке. Проведен эконометрический и кластерный анализ данных предприятия розничной торговли, проанализированы показатели и отмечены места, требующие пристального внимания компании. Разра-

ботано візуальна-аналітычыскае прыложэнне, якое дазваляе прааналізіраваць асновныя аналітычыскія паказателі прадпрыятця і прадпрыняць саотвештвуючыя управленчыскія рашэння, на аснованні гэтых даных.

Область применения: рознічная торгавля.

РЭЗІЮМЭ

Літвін Вераніка Валер'еўна

Матэматычныя метады і інструментальныя сродкі аналізу дадзеных прадпрыемствы рознічнага гандлю

Ключавыя словы: інтэлектуальны аналіз, мадэлі інтэлектуальнага аналізу, задачы інтэлектуальнага аналізу, выманне ведаў з баз дадзеных, візуалізацыя дадзеных, сродкі візуалізацыі.

Мэта работы: даследаванне матэматычных метадаў аналізу дадзеных для прадпрыемства рознічнага гандлю, а таксама выкарыстання інструментальных сродкаў для візуалізацыі прадастаўленых аналітычных дадзеных.

Атрыманыя вынікі і іх навізна: праведзена сістэматызацыя асноўных метадаў і задач інтэлектуальнага аналізу дадзеных прадпрыемствы рознічнага гандлю. Разгледжаны спосабы і сродкі візуалізацыі аналітычных дадзеных прадпрыемства. Структураваныя дадзеныя па ўсіх сродках візуалізацыі на сучасным рынку. Праведзены эканаметрычнага і кластарны аналіз дадзеных прадпрыемствы рознічнага гандлю, прааналізаваны паказчыкі і адзначаны месцы, якія патрабуюць пільнай увагі кампаніі. Распрацавана візуальна-аналітычнае дадатак, якое дазваляе прааналізаваць асноўныя аналітычныя паказчыкі прадпрыемства і зрабіць адпаведныя кіраўніцкія рашэнні, на падставе гэтых дадзеных.

Вобласць ужывання: рознічны гандаль.

SUMMARY

Litvin Veronika Valer'evna

Mathematical methods and tools for data analysis of retail enterprises

Keywords: Data Mining, models of Data Mining, tasks of Data Mining, Knowledge Discovery in Databases, data visualization, visualization tools.

The purpose of the work: research of mathematical methods of data analysis for a retail enterprise, and use of tools for visualization of provided analytical data.

The results and novelty: the systematization of the main methods and objectives of data mining to retailers. The methods and analytical data renderer enterprises. Structured data for all visualization tool on the market today. Conducted econometric analysis and cluster data to retailers, analyzed the figures and noted places that require attention of the company. Developed visual-analytic application that enables you to analyze the main analytical performance businesses and to take appropriate management decisions on the basis of these data.

Sphere of application: retail.