

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ

Кафедра экономической информатики

ГРЕБНЕВ

Роман Александрович

**Анализ поведения потребителей на рынке услуг телекоммуникаций
путем применения современных алгоритмов интеллектуального анализа
данных**

Магистерская диссертация

специальность 1-25 81 10 «Экономическая информатика»

Научный руководитель:
кандидат экономических наук,
доцент
Минюкович Екатерина
Александровна

Допущен к защите

«___» _____ 2019 г.

Зав. кафедрой экономической информатики

_____ Д.А. Марушко

кандидат экономических наук, доцент

Минск, 2019

ОГЛАВЛЕНИЕ

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ	4
ANNOTATION	6
АГУЛЬНАЯ ХАРАКТАРЫСТЫКА ПРАЦЫ	8
ВВЕДЕНИЕ	10
ГЛАВА 1 АНАЛИЗ ТЕНДЕНЦИЙ РАЗВИТИЯ СФЕРЫ ТЕЛЕКОММУНИКАЦИЙ	12
1.1 Условия развития индустрии телекоммуникаций в 21 веке: ретроспективный анализ и современные тренды развития	14
1.2 Бизнес-контексты применения данных в индустрии телекоммуникаций	19
1.2.1 Аналитика больших данных	21
1.2.2 Машинное обучение и искусственный интеллект	24
1.2.3 Интернет вещей	26
1.2.4 Роботизированная автоматизация процессов	27
1.3 Инновационные конкурентные стратегии в индустрии телекоммуникаций	29
1.3.1 Контентная стратегия	30
1.3.2 Предоставление data-продуктов	31
1.3.3 Финансовые продукты	32
Выводы по главе 1	34
ГЛАВА 2 АНАЛИТИЧЕСКИЕ ЭКОСИСТЕМЫ ТЕЛЕКОММУНИКАЦИОННЫХ КОМПАНИЙ	36
2.1 Анализ методологических подходов управления данными в телекоммуникационных компаниях	36
2.1.1 Методологии управления данными	37
2.1.2 Аналитическая культура и организационные структуры	42
2.1.3 Итерационные методологии анализа данных	46
2.2 Классификация типов данных телекоммуникационных компаний	52
2.2.1 Данные о клиентах	55
2.2.2 Данные об использовании услуг	55
2.2.3 Гео-данные и гео-информационные системы в индустрии телекоммуникаций	57
2.2.4 Информация из каналов заботы о клиенте	59
2.2.5 Данные о взаимодействии клиентов	60
2.3 Задачи анализа поведения потребителей в сфере телекоммуникаций	61
Выводы по главе 2	67
ГЛАВА 3 ВНЕДРЕНИЕ ПРОЦЕДУРЫ СКОРИНГА КЛИЕНТОВ МЕДИА-СЕРВИСА ДЛЯ КОМПАНИИ velcom A1	70
3.1 Анализ эффективности процедуры скоринга медиа-сервиса и бизнес-анализ	70
3.1.1 Описание текущей процедуры скоринга	71

3.1.2 Обоснование необходимости реинжиниринга бизнес-процессов скоринга	71
3.1.3 Критерии успешности проекта и требования к новой процедуре скоринга	73
3.1.4 Риски реализации проекта	74
3.1.5 Предварительный анализ	75
3.2 Оптимизация процедуры скоринга клиентов медиа-сервиса	77
3.2.1 Регрессия Кокса	78
3.2.2 Описание процедуры построения регрессии Кокса для скоринга клиентов медиа-сервиса	82
3.2.3 Описание новой процедуры построения скоринга	85
3.3 Анализ результатов внедрения новой процедуры скоринга	88
Выводы по главе 3	94
ЗАКЛЮЧЕНИЕ	96
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	98

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Объём работы: 96 страниц, 28 рисунков, 10 таблиц, 64 использованных источника.

Ключевые слова: ТЕЛЕКОММУНИКАЦИИ, ЦИФРОВАЯ ТРАНСФОРМАЦИЯ, СКОРИНГ, АНАЛИТИКА, БОЛЬШИЕ ДАННЫЕ, МОДЕЛЬ РЕГРЕССИИ КОКСА, ОПТИМИЗАЦИЯ, БИЗНЕС-ПРОЦЕССЫ, МАШИННОЕ ОБУЧЕНИЕ, МЕДИА-СЕРВИСЫ.

Объект исследования: аналитические экосистемы в телекоммуникационных компаниях.

Цель работы: Провести анализ трансформационных процессов на рынке телекоммуникаций; Повысить эффективность аналитической экосистемы телекоммуникационной компании velcom | A1 посредством изучения и внедрения лучших практик управления экосистемой данных на примере построение процедуры скоринга клиентов медиа-сервиса.

Методология исследования: анализ научной и научно-популярной литературы, статистических данных; количественный и качественный анализ; группировка экономической информации; общие приемы и законы логики (анализ, синтез, систематизация и др.); сравнительно-экономический метод, структурно-системный метод; математическое моделирование.

Научная новизна исследования: Современность и научная новизна данной работы заключается в том, что на текущий момент практики управления аналитической экосистемой данных в телекоммуникационных компаниях с учетом отраслевой специфики являются мало изученными. Проблематика качества цифровой трансформации белорусских телекоммуникационных компаний является одним из важнейших факторов, оказывающих влияние на цифровизационные процессы в масштабе страны. Также важно отметить, что развитие телекоммуникационной инфраструктуры является приоритетной задачей, отмеченной в пункте 6.1 программы социально-экономического развития Республики Беларусь на 2016 – 2020 гг.

В процессе работы были получены следующие результаты: раскрыта сущность трансформационных процессов, происходящих на мировом телекоммуникационном рынке; проанализированы методологические подходы к управлению данными в телекоммуникационных компаниях и выявлены наиболее эффективные практики; рассмотрены различные аспекты использования данных в коммерческих целях и определены точки роста компании velcom | A1 за счет внедрения подходов углубленной аналитики данных; разработано и внедрено аналитическое решение по сегментации клиентов медиа-сервиса для компании velcom | A1.

На основании выделенных лучших практик управления ценностью данных, в частности шестиэтапного аналитического цикла CRISP-DM, был успешно реализован аналитический проект по внедрению новой процедуры скоринга клиентов медиа-сервиса в компании velcom | A1.

Результаты исследования докладывались и обсуждались на научной конференции «76-я научная конференция студентов и аспирантов БГУ». Тезисы доклада приняты к публикации в материалах конференции.

ANNOTATION

Scope of work: 96 pages, 28 figures, 10 tables, 64 sources used.

Key words: TELECOMMUNICATIONS, DIGITAL TRANSFORMATION, SCORING, ANALYTICS, BIG DATA, COX REGRESSION, OPTIMIZATION, BUSINESS PROCESSES, MACHINE LEARNING, MEDIA SERVICES.

Subject of study: analytical ecosystem for telecommunication companies.

Purpose of work: To analyze the transformation processes in the telecommunications market; in order to increase the efficiency of the analytical ecosystem of the telecommunication company velcom | A1 through the study and implementation of best practices of data ecosystem management on the example of building a scoring procedure for media service customers.

Research methodology: analysis of scientific and popular science literature, statistical data; quantitative and qualitative analysis; grouping of economic information; general techniques and laws of logic (analysis, synthesis, systematization, etc.); comparative-economic method, structural-system method; mathematical modeling.

Scientific novelty of the research: The modernity and scientific novelty of this work lies in the fact that at the moment the practice of management of the analytical ecosystem in telecommunication companies, taking into account industry specifics, is little studied. The problem of the quality of digital transformation of Belarusian telecommunications companies is one of the most important factors influencing the digitalization processes in the country. It is also important to note that the development of telecommunications infrastructure is a priority, as stated in paragraph 6.1 of the program of socio-economic development of the Republic of Belarus for 2016 – 2020.

The following results were obtained: the essence of the transformation processes taking place in the global telecommunications market was revealed; methodological approaches to data management in telecommunications companies were analyzed and the most effective practices were identified; various aspects of data use for commercial purposes were considered and the growth points of velcom | A1 were determined through the introduction of in-depth data analytics approaches; an analytical solution for segmentation of media service clients for velcom | A1 was developed and implemented.

On the basis of the selected best practices of data value management, in particular the six-stage analytical cycle CRISP-DM, a new procedure for scoring media service customers in the company velcom | A1 was successfully implemented as an analytical project.

The results of the research were presented and discussed at the scientific conference "76th scientific conference of students and postgraduates of BSU". Abstracts of the report are accepted for publication in the conference proceedings.

Отчет о проверке на заимствования №1



Автор: r.grebnev14@gmail.com / ID: 6832763

Проверяющий: r.grebnev14@gmail.com / ID: 6832763

Отчет предоставлен сервисом «Антиплагиат»- <http://users.antiplagiat.ru>

ИНФОРМАЦИЯ О ДОКУМЕНТЕ

№ документа: 1
 Начало загрузки: 04.06.2019 22:25:11
 Длительность загрузки: 00:00:04
 Имя исходного файла: Магистерская-работа-Гребнев
 Размер текста: 2712 кБ
 Символов в тексте: 198846
 Слов в тексте: 22882
 Число предложений: 1038

ИНФОРМАЦИЯ ОБ ОТЧЕТЕ

Последний готовый отчет (ред.)
 Начало проверки: 04.06.2019 22:25:15
 Длительность проверки: 00:00:07
 Комментарии: не указано
 Модули поиска: Модуль поиска Интернет



ЗАИМСТВОВАНИЯ 4,03% ЦИТИРОВАНИЯ 0% ОРИГИНАЛЬНОСТЬ 95,97%

Заимствования — доля всех найденных текстовых пересечений, за исключением тех, которые система отнесла к цитированиям, по отношению к общему объему документа.
 Цитирования — доля текстовых пересечений, которые не являются авторскими, но система посчитала их использование корректным, по отношению к общему объему документа. Сюда относятся оформленные по ГОСТу цитаты; общепотребительные выражения; фрагменты текста, найденные в источниках из коллекций нормативно-правовой документации.
 Текстовое пересечение — фрагмент текста проверяемого документа, совпадающий или почти совпадающий с фрагментом текста источника.
 Источник — документ, проиндексированный в системе и содержащийся в модуле поиска, по которому проводится проверка.
 Оригинальность — доля фрагментов текста проверяемого документа, не обнаруженных ни в одном источнике, по которым шла проверка, по отношению к общему объему документа.
 Заимствования, цитирования и оригинальность являются отдельными показателями и в сумме дают 100%, что соответствует всему тексту проверяемого документа.
 Обращаем Ваше внимание, что система находит текстовые пересечения проверяемого документа с проиндексированными в системе текстовыми источниками. При этом система является вспомогательным инструментом, определение корректности и правомерности заимствований или цитирований, а также авторства текстовых фрагментов проверяемого документа остается в компетенции проверяющего.

№	Доля в отчете	Источник	Ссылка	Актуален на	Модуль поиска
[01]	2,94%	Применение регрессии Кокса в здравоохранении с использованием пакета статистических програ...	https://cyberleninka.ru	07 Ноя 2018	Модуль поиска Интернет
[02]	0,21%	https://www.mckinsey.com/sitemap.xml (1/3)	https://mckinsey.com	26 Мар 2018	Модуль поиска Интернет
[03]	0,18%	Применение RFM-анализа в сегментировании клиентской базы / Блог компании UniSender / Хабра...	http://habrahabr.ru	раньше 2011	Модуль поиска Интернет

Еще источников: 12
 Еще заимствований: 0,7%

АГУЛЬНАЯ ХАРАКТАРЫСТЫКА ПРАЦЫ

Аб'ём працы: 96 старонак, 28 малюнкаў, 10 табліц, 64 выкарыстаных крыніцы.

Ключавыя словы: ТЭЛЕКАМУНІКАЦЫІ, ЛІЧБАВАЯ ТРАНСФАРМАЦЫЯ, СКОРЫНГ, АНАЛІТЫКА, ВЯЛІКІЯ ДАДЗЕННЫЯ, МАДЭЛЬ РЭГРЭСІІ КОКСУ, АПТЫМІЗАЦЫЯ, БІЗНЕС-ПРАЦЭСЫ, МАШЫННАЕ НАВУЧАННЕ, МЕДЫЯ-СЭРВІСЫ.

Аб'ект даследавання: аналітычныя якасці тэлекамунікацыйных кампаніяў.

Мэта працы: Правесці аналіз трансфармацыйных працэсаў на рынку тэлекамунікацый; павысіць эфектыўнасць аналітычнай якасці тэлекамунікацыйнай кампаніі velcom | A1 з дапамогай вывучэння і ўкаранення лепшых практык кіравання якасці дадзеных на прыкладзе пабудовы працэдуры скорынку кліентаў медыя-сэрвісу.

Метадалогія даследавання: аналіз навуковай і навукова-папулярнай літаратуры, статыстычных дадзеных; колькасны і якасны аналіз; групоўка эканамічнай інфармацыі; агульныя прыёмы і законы логікі (аналіз, сінтэз, сістэматызацыя і інш.); параўнальна-эканамічны метады, структурна-сістэмны метады; матэматычнае мадэляванне.

Навуковая навізна даследавання: Сучаснасць і навуковая навізна дадзенай працы заключаецца ў тым, што на бягучы момант практыкі кіравання аналітычнай якасці дадзеных у тэлекамунікацыйных кампаніяў з улікам галіновай спецыфікі з'яўляюцца мала вывучанымі. Праблематыка якасці лічбавай трансфармацыі беларускіх тэлекамунікацыйных кампаніяў з'яўляецца адным з найважнейшых фактараў, якія аказваюць уплыў на цыфравізацыйныя працэсы ў маштабе краіны. Таксама важна адзначыць, што развіццё тэлекамунікацыйнай інфраструктуры з'яўляецца прыярытэтнай задачай, адзначанай у пункце 6.1 Праграмы сацыяльна-эканамічнага развіцця Рэспублікі Беларусь на 2016-2020 гг.

У працэсе работы былі атрыманы наступныя вынікі: раскрытая сутнасць трансфармацыйных працэсаў, якія адбываюцца на сусветным тэлекамунікацыйным рынку; прааналізаваны метадалагічныя падыходы да кіравання дадзенымі ў тэлекамунікацыйных кампаніяў і выяўлены найбольш эфектыўныя практыкі; разгледжаны розныя аспекты выкарыстання даных у камерцыйных мэтах і вызначаны пункты росту кампаніі velcom | A1 за кошт укаранення падыходаў паглыбленай аналітыкі дадзеных; распрацавана і ўкаранёна аналітычнае рашэнне па сегментацыі кліентаў медыя-сэрвісу для кампаніі velcom | A1.

На падставе вылучаных лепшых практык кіравання каштоўнасцю дадзеных, у прыватнасці шасціэтапнага аналітычнага цыкла CRISP-DM, быў паспяхова рэалізаваны аналітычны праект па ўкараненні новай працэдуры скорынку кліентаў медыя-сэрвісу ў кампаніі velcom | A1.

Вынікі даследавання дакладваліся і абмяркоўваліся на навуковай канферэнцыі «76-я навуковая канферэнцыя студэнтаў і аспірантаў БДУ». Тэзісы дакладу прыняты да публікацыі ў матэрыялах канферэнцыі.

ВВЕДЕНИЕ

Динамика развития индустрии телекоммуникаций в 21 веке оказывает влияние на все секторы экономики, а также определяет их продуктивность. Анализ тенденций развития современного телекоммуникационного сектора позволяет говорить об интенсификации процессов цифровой трансформации компаний и их переходе к бизнес-моделям, основанным на данных. Внедрение подходов обработки больших массивов данных является ключевой конкурентной стратегией компаний мирового рынка телекоммуникаций, направленной на компенсацию эффекта коммодизации услуг связи.

Массовая переориентация на данные как на источник конкурентного преимущества, а также становление данных одним из ключевых активов телекоммуникационных компаний обуславливает необходимость детального изучения потенциала данных, которыми обладают телекоммуникационные компании будучи провайдерами сетевой инфраструктуры. Важным условием результативной реализации этого потенциала является построение эффективно функционирующей экосистемы управления данными, в частности аналитической экосистемы.

На примере реализации проекта по внедрению аналитического решения в рамках контентной стратегии компании velcom | A1, рассмотрены подходы к извлечению ценности из данных, а также методы монетизации данных.

Цель данной работы можно сформулировать следующим образом: Провести анализ трансформационных процессов на рынке телекоммуникаций; Повысить эффективность аналитической экосистемы телекоммуникационной компании velcom | A1 посредством изучения и внедрения лучших практик управления экосистемой данных на примере построение процедуры скоринга клиентов медиа-сервиса.

Для достижения цели, были поставлены и решены следующие задачи:

1. Раскрыть сущность трансформационных процессов в сфере телекоммуникаций, происходящих в результате внедрения технологий, связанных с анализом больших данных.
2. Проанализировать методологические подходы к управлению данными в телекоммуникационных компаниях и выявить наиболее эффективные практики.
3. Рассмотреть различные аспекты использования данных в коммерческих целях и определить возможные точки роста бизнеса за счет внедрения подходов углубленной аналитики данных.
4. Разработать и внедрить аналитическое решение по сегментации клиентов медиа-сервиса на базе изученных практик для компании velcom | A1.

Современность и научная новизна данной работы заключается в том, что на текущий момент практики управления аналитической экосистемой данных в телекоммуникационных компаниях с учетом отраслевой специфики являются мало изученными. Проблематика качества цифровой трансформации белорусских телекоммуникационных компаний является одним из важнейших факторов, оказывающих влияние на цифровизационные процессы в масштабе страны. Также важно отметить, что развитие телекоммуникационной инфраструктуры является приоритетной задачей, отмеченной в пункте 6.1 программы социально-экономического развития Республики Беларусь на 2016 – 2020 гг.

Развитие сферы телекоммуникаций тесно связано с эффективностью экономических субъектов и ростом национального благосостояния. Поэтому следует отметить особую важность рассмотрения факторов и механизмов трансформации телекоммуникационной сферы, так как в силу высокой системной значимости результатов работы телекоммуникационного сектора, возникает существенный мультипликативный эффект, воздействующий на всю экономику.

Объект исследования – аналитические экосистемы в телекоммуникационных компаниях.

Предметом исследования является применение аналитики больших данных в целях оптимизации бизнес-процессов в телекоммуникационных компаниях.

ГЛАВА 1

АНАЛИЗ ТЕНДЕНЦИЙ РАЗВИТИЯ СФЕРЫ ТЕЛЕКОММУНИКАЦИЙ

Технологическое развитие в 21 веке определяет структуру общественных отношений, глубоко трансформируя различные аспекты человеческой жизнедеятельности. Технократизация социо-экономических институтов приводит к реструктурированию общественных связей, а также перераспределению капитала в наукоемкие сферы. Постепенное превращение информации в важнейший фактор производства демонстрирует необходимость изучения проблематики менеджмента этого ресурса, а также оценки путей его эффективного использования. Развитие сферы телекоммуникаций тесно сопряжено с процессами информатизации общества, увеличением качества жизни, а также с развитием цифровой экономики. Телекоммуникационная инфраструктура выполняет ключевые функции по передаче цифрового ресурса, и является одним из основных факторов, обуславливающим динамику современных цифровизационных процессов.

На данный момент наблюдается высвобождение накопленной критической массы технологических инноваций, а также повышение степени готовности различных экономических сфер к качественно новому этапу развития, драйвером которого являются данные и способы их эффективного использования.

Телекоммуникационные компании не только активно участвуют в цифровой трансформации экономики, но и сами подвергаются значительным цифровизационным изменениям, которые коренным образом меняют ландшафт сферы телекоммуникаций, тем самым определяя более масштабные кросс-индустриальные процессы создания коннективной ценности. Сфера телекоммуникаций оказывает положительный экстернальный эффект на множество экономических сфер в связи с интенсификацией взаимодействия экономических агентов, повышения качества этого взаимодействия, а также усложнения организационных структур.

В исследовании центра европейских экономических исследований «Обзор влияния телекоммуникационных сетей и широкополосного доступа в Интернет» были сделаны следующие выводы касательно степени интенсивности и направления влияния развития сферы телекоммуникаций на экономику [1]:

- На макроэкономическом уровне развертывание широкополосной связи не только приводит к более высоким темпам экономического роста и производительности, но также помогает создавать новые рабочие места и способствует экономическому развитию в сельской местности.

- Микроэкономические эффекты, оказываемые развертыванием широкополосной связи, дифференцированы и приводят к неоднородным экономическим эффектам. Высококвалифицированные рабочие и высокотехнологичные компании получают преимущество в реализации своего потенциала, в то время как сравнительно менее квалифицированные рабочие и компании с низкой наукоемкостью не являются бенефициарами повышенной коннективности.
- Все рассмотренные в обзоре исследования, посвященные эффектам от внедрения как проводной, так и беспроводной связи, показывают существенное положительное влияние на производительность отдельных фирм, экономических секторов, а также экономических систем в целом.

Таким образом, развитие сферы телекоммуникаций тесно связано с эффективностью экономических субъектов и ростом национального благосостояния. Поэтому следует отметить особую важность рассмотрения факторов и механизмов трансформации телекоммуникационной сферы, так как в силу высокой системной значимости результатов работы телекоммуникационного сектора, возникает существенный мультипликативный эффект на всю экономику.

Также актуальность изучения факторов развития телекоммуникационного рынка обусловлена масштабом индустрии, которую можно продемонстрировать на основании статистики ETNO [2]:

- **778М** абонентов мобильной связи в странах Европы;
- Уровень проникновения мобильной связи в Европе достигает **126%**;
- В сфере телекоммуникаций в Европе занято более **650 000** человек;
- Совокупный доход европейских телекоммуникационных компаний оценивается в **260 млрд. евро** на конец 2017 года;
- Средний темп роста годовой валовой выручки в индустрии составил **0.3%** в 2017 г.;
- Телекоммуникации составили **26.7%** европейского ИКТ-рынка в 2017 г.;

Для понимания того, какие качественные изменения происходят в ключевой для информатизации общества сфере телекоммуникаций необходимо рассмотреть историческую динамику развития мирового рынка телекоммуникаций, технологические концепции, меняющие телекоммуникационную сферу, а также стратегии адаптации компаний рынка к новым экономическим условиям.

Так проблемы качества и доступности услуг связи по мере коммодизации телекоммуникационных услуг отступают на второй план в массовом восприятии потребителей. Тем не менее инфраструктура передачи данных, являющаяся базовой для многих цифровых процессов, первична и играет критическую роль в цепочке создания ценности в информационной экономике. Экономические агенты в большей степени осознают ценность

информации, в связи с чем возникают новые индустрии. Например, на стыке промышленности и информационных технологий появилась индустрия интернета вещей. Накопление данных, а также наращивание вычислительных мощностей в течение последнего десятилетия привело к развитию концепции больших данных и совершенствованию подходов управления данными. Также синергетический эффект от соприкосновения сферы информационных технологий и различных прикладных сфер активизировал развитие подходов, ассоциируемых с искусственным интеллектом и машинным обучением. Помимо этого, наблюдается активное слияние телекоммуникаций с другими отраслями, такими как: медиа, информационные технологии, а также финансовые услуги [3].

Телекоммуникационные сети, обеспечивающие циркуляцию данных, можно сравнить с нервной системой биологического организма, которая является разветвленной взаимосвязанной сложной системой, регулирующей его работу и реагирующей на внутренние и внешние изменения [4]. Аналогично актуальные оперативные данные питают множество индустрий, создавая петлю обратной связи в организациях, что поддерживает механизм реагирования на динамично меняющиеся экономические условия.

Большинство сфер жизнедеятельности современного общества переориентируются на интенсивное создание, обмен и утилизацию данных. Проблеме данных уделяется не меньшее внимание, чем энергоресурсам. Данные сравниваются с новой нефтью в связи с их возрастающей ролью в экономических процессах: они выполняют координирующую функцию в обществе, применяются в процессах принятия решений, а также лежат в основе цепочек создания ценности современного высокотехнологичного бизнеса.

Опираясь на вышесказанное можно констатировать: функционал, надстраиваемый над физическими сетями передачи данных приводит к глубокой трансформации общества и экономических процессов, меняет принципы взаимодействия и формирует новые цепочки создания ценности. В связи с этим требуют проработки вопросы максимально эффективного использования существующей инфраструктуры, извлечения дополнительной ценности из данных, накапливаемых компаниями на этом рынке и выделение точек роста для обеспечения глобального позитивного экономического эффекта.

1.1 Условия развития индустрии телекоммуникаций в 21 веке: ретроспективный анализ и современные тренды развития

Индустрия телекоммуникаций как индустрия, выполняющая функцию администрирования разветвленных каналов передачи данных в обществе, претерпевает существенные изменения в связи с динамичным развитием информационно-интенсивных экономических сфер, а также развитием сопряженных технологических концепций. Однако для определения

ключевых характеристик развития индустрии телекоммуникаций также необходимо учитывать исторический контекст развития этой сферы, в связи с чем актуализируется задача ретроспективного анализа. Именно сочетание исторических и новых факторов, формирующих индустрию телекоммуникаций, определяет текущие методы конкуренции, а также ее перспективы в целом.

В связи с отсутствием конвенциональных подходов к периодизации развития современной телекоммуникационной сферы, представляется целесообразным выделить четыре этапа развития рынка телекоммуникационных услуг, отличающихся в различных странах в зависимости от степени развитости национального сектора телекоммуникаций:

1. Первый этап - зарождение мирового рынка телекоммуникаций, который продолжался до середины 90-х годов XX века. Этот этап характеризуется умеренными темпами роста, что связано со сравнительно небольшим спросом на телекоммуникационные услуги и явным преобладанием монополистических структур. В фазе зарождения рынок демонстрировал высокую гетерогенность и отсутствие единых технологических подходов [5];
2. Второй этап развития телекоммуникационных услуг продолжался с конца 1990-х до начала 2010-х. Он характеризуется бурным развитием индустрии, существенным повышением спроса на услуги и развитием технологической базы;
3. В странах, в которых телекоммуникационный рынок пришел к точке насыщения, можно выделить третий этап, который продолжался с 2012 года по 2015 год. Для него характерно замедление темпов роста в терминах количества клиентов и выручки, насыщение рынка и достижение проникновения услуг телекоммуникаций уровня близкого к 100% [6]. Также на данном этапе можно отметить исчерпание консервативных источников роста, ожесточение конкурентной борьбы и падение средней маржинальности сферы.
4. Четвертый этап развития рынка начался в 2015 году и продолжается в настоящий момент. Переход от 3 к 4 этапу происходит вместе с глубокой трансформацией отрасли под воздействием внедряемых инновационных ИКТ-решений, реинжиниринга бизнес-процессов в компаниях отрасли, развития смежных сфер, задействующих инфраструктуру телекоммуникационных операторов и развития конвергентных предложений для клиентов.

Ниже представлена авторская периодизация развития рынка телекоммуникаций (Рисунок 1.1).



Рисунок 1.1 – Этапы развития современного мирового рынка телекоммуникаций

Примечание – Разработка автора

Для последующего анализа тенденций развития индустрии телекоммуникаций наиболее актуальными являются эффекты, представленные на втором, третьем и четвертом этапах. Динамика развития индустрии на втором этапе резко контрастирует со сдержанным ростом, демонстрируемым мировым рынком телекоммуникаций на третьем этапе. Четвертый этап развития, начавшийся в 2015 году, демонстрирует осознание компаниями-представителями индустрии новых конъюнктурных условий и их стремление получить максимальный выигрыш от внедрения технологических инноваций. Также эти этапы развития отличаются конкурентными стратегиями, применяемыми компаниями.

На втором этапе происходило активное формирование рынка телекоммуникаций, в этот период данная сфера представляла собой одну из самых динамично развивающихся отраслей экономики. Растущая потребность общества в качественной связи сопровождалась укоренением услуг связи в различных сферах жизнедеятельности. Также росло количество индивидуальных средств связи, что обеспечивало лавинообразное развитие сферы: наблюдалось развитие сетей передачи данных, изначально дорогие технологические решения оптимизировались. Помимо этого, происходил переход к использованию ограниченного круга наиболее современных стандартов связи. Стоимость как терминального оборудования, так и услуг связи постепенно снижалась. На фоне бурного роста мирового рынка телекоммуникаций происходила либерализация деятельности телеком-компаний, повышалась открытость национальных рынков, а также наблюдался активный процесс создания совместных предприятий с участием иностранных операторов. Деятельность телекоммуникационных компаний диверсифицировалась, что сопровождалось расширением спектра предоставляемых услуг. Нарастающая конкуренция между

телекоммуникационными компаниями привела к интенсификации слияний и поглощений, что в результате вылилось в концентрацию капитала в руках крупнейших представителей рынка [2].

Третий этап отражает проблему деактуализации бизнес-моделей телеком-компаний и подчеркивает необходимость пересмотра формулы лидерства на рынке, согласно которой владение сетевой инфраструктурой и наличие обширной дистрибьюторской сети являлись основными конкурентными преимуществами. Жесткая ценовая конкуренция, связанная с коммодизацией предоставляемых услуг фиксированной и голосовой связи привела к снижению выручки компаний, что отражено на графике (Рисунок 1.2) [7].

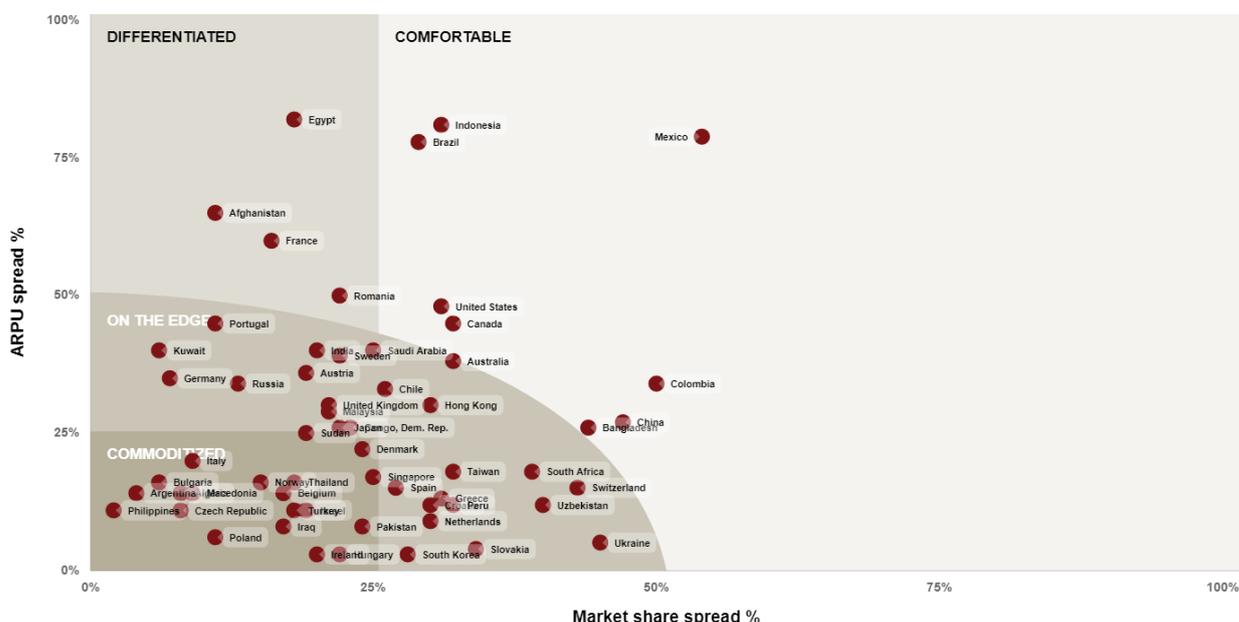


Рисунок 1.2 – Индекс коммодизации услуг связи

Примечание

—

Источник:

PwC

(<https://www.strategyand.pwc.com/wirelesscommoditization>)

Индекс коммодизации, учитывающий степень дифференциации предложений на рынке и рассчитываемый на ежегодной основе компанией PwC, повысился в Европе и Северной Америке на 29% и 39% соответственно за период с 2007 по 2017 год [7]. По данным консалтинговой компании McKinsey средние доходы в индустрии снижались в среднем на 6 процентов в год с 2010 по 2016 годы. [ссылка]. В своем исследовании компания PwC показала, что средний чек на одного клиента в телеком-индустрии снизился с 2016 по 2017 год на 3%, упав при этом за 10 лет с 2007 по 2017 годы на 34%. На графике ниже представлена динамика сервисной выручки телекоммуникационных операторов с 2012 по 2017 годы по данным европейской ассоциации операторов телекоммуникационных сетей ETNO (Рисунок 1.3).

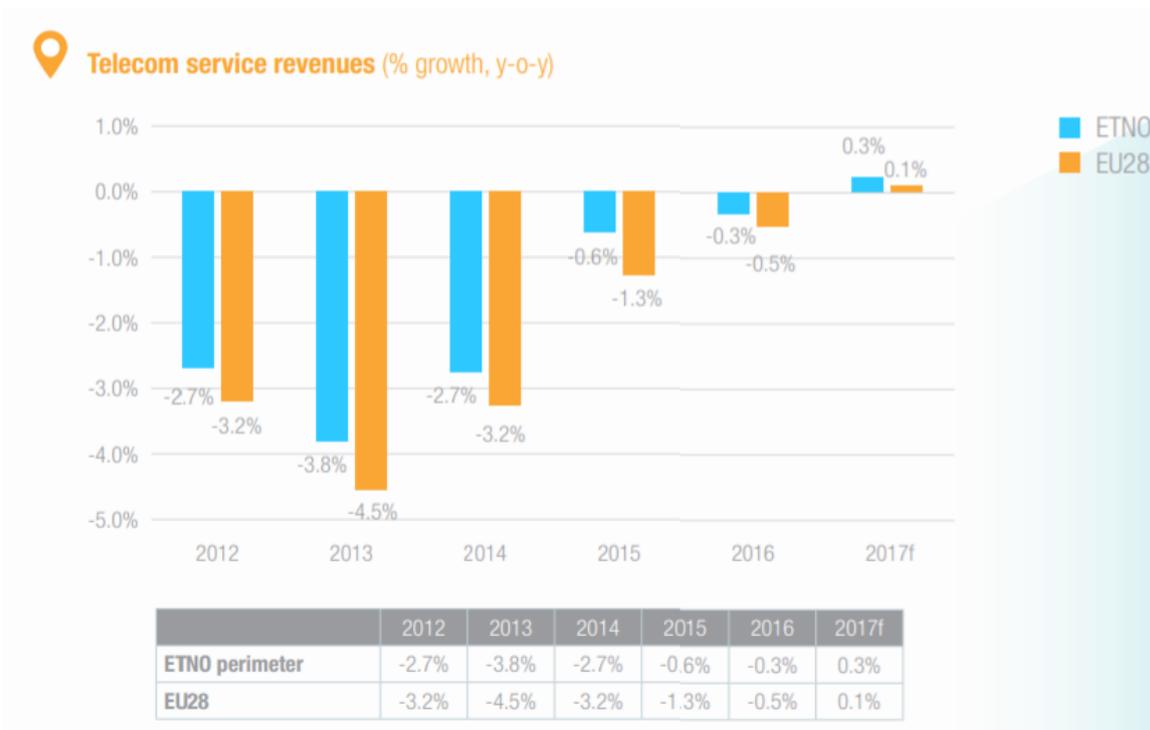


Рисунок 1.3 – Динамика сервисной выручки телекоммуникационных операторов

Примечание – Источник: ETNO (<https://etno.eu/datas/publications/annual-reports>)

Удешевление и снижение потребления голосового трафика, а также связанное с этим падение выручки отчасти компенсировалось стремительным ростом потребления интернет-трафика. Тем не менее, по мнению экспертов McKinsey острая ценовая конкуренция и борьба за долю рынка в долгосрочном периоде усугубит положение в отрасли, оказывая серьезное ценовое давление на компании. В связи с этим рынок телекоммуникаций сталкивается с задачей поиска новых дифференцирующих факторов и новых конкурентных преимуществ.

Отличительной чертой четвертого этапа развития рынка телекоммуникаций является создание новых ценностных предложений для клиентов, а также повышение дифференциации продукта, за счет формирования уникального клиентского опыта. Современные телекоммуникационные компании начинают осознавать важность создания целостной продуктовой экосистемы, создающей эффект “бесшовного” использования, что в терминах институциональной экономики синонимично созданию среды с более низкими транзакционными издержками (оплата всех сервисов через один пользовательский интерфейс, интеграция различных продуктов в единую экосистему и др.). Также стоит отметить, что компании стремятся к созданию конвергентных предложений, путем расширения продуктовой линейки через вертикальную и горизонтальную интеграцию. Важнейшей особенностью четвертой фазы является становление сферы телекоммуникаций точкой взаимодействия и кросс-фасилитации различных технологий, в связи с чем многие телеком-операторы становятся

интеграторами технологических решений (интернет вещей, умная городская инфраструктура, промышленный интернет и т. д.). Этот процесс является двунаправленным и взаимовыгодным: телеком-компании получают конкурентное преимущество по мере интеграции решений по аналитике больших данных, применению машинного обучения в целях построения интеллектуальных систем принятия решений, внедрения инструментов роботизированной автоматизации процессов. Для четвертой фазы характерно превращение телеком-оператора из оператора услуг связи в оператора сервисов посредством применения инновационных конкурентных стратегий и внедрения современных технологических концепций в бизнес-процессы.

Таким образом, центральной идеей, доминирующей на четвертой фазе развития рынка телекоммуникаций являются данные, а также компоненты информационной инфраструктуры, генерирующие потоки данных, и способы извлечения ценности из данных.

Эволюция сферы телекоммуникаций подчиняется циклу Gartner, однако нарастающая динамика отрасли на 4 фазе ее развития показывает перспективы выхода с плато продуктивности на новую итерацию цикла за счет создания нового класса ценности на рынке телекоммуникаций [8].

Для выделения наиболее перспективных направлений развития телекоммуникационных компаний необходимо выделить лучшие практики на основании изучения международного опыта и детальный анализ трансформационного потенциала современных технологических концепций для индустрии телекоммуникаций.

1.2 Бизнес-контексты применения данных в индустрии телекоммуникаций

Для экономического развития в 21 веке характерны интенсификация транзакционной активности, снижение транзакционных издержек, а также появление институтов нового образца, в основе которых лежат системы управления данными. Увеличение количества сделок, совершаемых в экономике, и углубление уровня взаимодействия различных контрагентов приводит к созданию информационного следа, который может быть использован для принятия экономическими агентами решений, основанных на данных.

Цифровая экономика как информационная инфраструктура аналоговой экономики, помимо физической инфраструктуры, обеспечиваемой и администрируемой индустрией телекоммуникаций, формируется из множества технологических подходов, активно используемых в отраслевых цепочках создания ценности информационной экономики. Среди них можно выделить подходы, которые наиболее тесно взаимодействуют со сферой телекоммуникаций. В своем отчете компания Ernst & Young выделяет 4 ключевых технологических подхода, определяющих будущее отрасли телекоммуникаций [9]:

- Аналитика больших данных (Big Data);
- Машинное обучение (Machine Learning) и искусственный интеллект (Artificial Intelligence);
- Интернет вещей (Internet of Things) и промышленный интернет (Industrial Internet);
- Роботизированная автоматизация процессов (Robotic Process Automation).

Доминирующей концепцией на четвертом этапе развития индустрии телекоммуникаций является управление данными. На высоко конкурентном рынке телекоммуникаций перед бизнесом возникает ряд задач, связанных с извлечением ценности из накапливаемого многообразия данных.

С развитием и интеграцией решений класса Business Intelligence телеком-компании начали накапливать значительные массивы данных о своей операционной деятельности. Крупные компании первыми столкнулись с проблемой избыточности структурированных и неструктурированных данных, используемых в целях администрирования бизнеса, в связи чем появилась потребность в систематизации и приоритезации решения задач управления данными. Вместе с накоплением данных, трансформировалось восприятие информационных активов предприятия. В среде менеджмента компаний произошел переход к осознанию того, что конкурентоспособность компаний определяется глубиной понимания внешних условий и внутренних бизнес-процессов. Естественный способ извлечения ценности из таких данных в - моделирование процессов и явлений, оказывающих прямой эффект на операционную деятельность компании.

Компании, у которых главной целью является максимизация прибыли в долгосрочном периоде, возникает потребность в тестировании бизнес-гипотез, касательно оптимальной стратегии развития. В связи с массовым внедрением аналитических систем, говорят о новой эволюционной ступени развития менеджмента - управлении на основе статистики.

Пионером внедрения подходов управления компанией на основе данных является американский ученый Уильям Эдвардс Деминг. Деминг, автор концепции статистического контроля качества, произвел настоящую революцию в японской промышленности, инициировав массовое внедрение подходов управления, базирующихся на статистике, значительно повысив качество продукции. Имплементация идей Деминга фактически превратила японскую промышленность в одну из наиболее конкурентоспособных в мире [10].

Заложенные Демингом в середине 20 века принципы управления не потеряли актуальности и в эпоху массовой цифровизации бизнеса. Вместе с бурным развитием технологий обработки данных и массовым применением статистического моделирования, а также с развитием методов машинного обучения и искусственного интеллекта, происходит переход к бизнес-моделям, на основе данных (от англ. data-driven approach), что актуально для текущего состояния индустрии телекоммуникаций.

Отмеченные ранее подходы участвуют в генерации данных, а также предлагают новые кросс-дисциплинарные методы утилизации данных для оптимизации бизнеса и поиска точек роста. В связи с этим необходимо детально рассмотреть влияние наиболее значимых технологий на развитие индустрии телекоммуникаций, так как именно они являются источником трансформационных процессов, наблюдаемых на четвертом этапе развития мирового рынка телекоммуникаций.

1.2.1 Аналитика больших данных

Телекоммуникационные компании собирают огромные объемы данных из записей о вызовах, использования мобильных телефонов, сетевого оборудования, журналов серверов, биллинговых систем и социальных сетей, генерируя многомерные массивы информации о своих клиентах и сетях. Современные бизнес-системы, используемые в телеком-компаниях, относятся к классу больших данных, так как обрабатывается детализированная информация о миллионах транзакций, совершаемых клиентами в рамках различных бизнес-измерений, определяемых предоставляемыми компаниями сервисами и продуктами. Важнейшим аспектом построения аналитической экосистемы являются данные, которые пользуются доверием у бизнес-пользователей и лиц, принимающих решения. Для этого необходимо выполнение требований достаточности уровня развития культуры данных, наличие необходимых элементов инфраструктуры данных и целостная система менеджмента данных. Все эти элементы являются базовыми для извлечения ценности из накапливаемых массивов информации. Вместе с развитием концепции больших данных происходит переход от восприятия представителями бизнеса данных как побочного продукта, создаваемого в процессе функционирования бизнеса, к их восприятию как главного и наиболее ценного нематериального актива компании, выступающего в роли ключевого рыночного дифференциального фактора. Аналитика больших данных по отчетам консалтинговой компании Deloitte является одной из наиболее востребованных технологических инноваций в 2018 году, так как именно большие данные как комплекс технологических решений позволяют сделать бизнес более структурированным, а также более эффективно управлять цепочкой создания ценности [11].

С ростом клиентской базы, расширением спектра предоставляемых сервисов, а также развитием аналитической культуры в сфере телекоммуникаций, многие компании начали сталкиваться с проблемой избыточности данных. Среди ключевых проблем, связанных с администрированием данных выделяют следующие: сбор, хранение, обработка и анализ данных.

Неэффективное управление большими данными может привести к повышению расходов, а также к снижению эффективности работы и

конкурентоспособности. И напротив, разумная стратегия по работе с большими данными помогает телекоммуникационным компаниям сократить расходы и получить дополнительные преимущества за счет повышения операционной эффективности.

Несмотря на новизну концепции больших данных (она впервые появилась в конце 2000-х), понимание больших данных успело претерпеть эволюцию, которую можно описать следующими этапами: описательный анализ (что произошло, какова причина произошедшего), прогностический анализ (оценка вероятности событий в будущем, предсказательные модели, интеллектуальный мониторинг), предписывающий анализ (формирование рекомендаций, контекстно-зависимая поддержка принятия решений).

Телекоммуникационные компании опираются на внутренние аналитические системы для принятия решений в изменчивой бизнес-среде. Аналитика данных используется в самых различных бизнес-контекстах: системы поддержки и принятия решений, маркетинговая аналитика, анализ поведенческого профиля клиентов, таргетинговая реклама, системы сопровождения, удержание клиентов и другие. Формализованные задачи, связанные с аналитикой данных, рассматриваются детально во второй главе.

Для понимания парадигмы больших данных и оказываемого ей влияния на сферу телекоммуникаций, следует дать ей определение. В научной и бизнес-средах наблюдается несогласованность в определении того, что именно следует понимать под большими данными.

Так, согласно определению Википедия, большие данные (от англ. Big Data) – обозначение структурированных и неструктурированных данных огромных объемов и значительного многообразия, эффективно обрабатываемых горизонтально масштабируемыми программными инструментами для решения задач класса Business Intelligence, альтернативными традиционным системам управления базами данных [12].

Согласно определению издательства Rusbase, “большие данные - это различные инструменты, подходы и методы обработки как структурированных, так и неструктурированных данных для использования их в целях решения конкретных бизнес-задач” [13]. В данном определении ключевой особенностью больших данных является способность извлекать ценность для бизнеса.

Компанией Gartner дается следующее определение: “Большие данные - это данные, которые характеризуются объемом, разнообразием и скоростью, с которой структурированные и неструктурированные данные поступают по сетям передачи в процессоры и хранилища, наряду с процессами преобразования этих данных в ценную для бизнеса информацию” [14]. В определении подчеркивается требование к высокой оперативности обрабатываемых данных, а также делается акцент на работу также с неструктурированными данными.

Оксфордский словарь определяет большие данные как наборы данных очень большого размера, которые могут быть проанализированы с

использованием больших вычислительных мощностей, для выявления трендов и ассоциативных правил, особенно, касательно поведения людей и их взаимодействия [15].

Сайт кембриджского университета дает несколько иное определение большим данным, отсылающее к центральной роли человека, но в неразрывной связке с технологиями: Очень большие массивы данных, производимые людьми в процессе пользования интернетом, которые могут быть сохранены, поняты и использованы только с помощью специальных инструментов и методов [16]. Тем не менее, данные, создаваемые в современных системах, нельзя редуцировать лишь к данным о человеческой жизнедеятельности, но в связи с возрастающей значимостью интернета вещей и развитием сенсорных систем, большие данные также генерируются и объединены в сети объектами инфраструктуры и роботизированными системами, в связи с чем определение необходимо дополнить разнообразными источниками создания данных, не зависящих от человека.

Исходя из вышеизложенных определений, представляется возможным определить большие данные в контексте сферы телекоммуникаций следующим образом:

Большие данные - это данные, характеризующиеся большим размером, разнообразной степенью качества и структурированности, генерируемые различными источниками - как в качестве результата информационной жизнедеятельности человека, так и автономными системами, а также требующими использования специальных методик обработки с целью извлечения ценности для бизнеса.

Слово «большие» в устоявшемся словосочетании «большие данные» означает скорее качественную характеристику, нежели количественную. Большие данные обычно определяют с точки зрения проблем управления данными, которые не удается решить в рамках традиционных баз данных в силу объема, разнообразия данных и требований к скорости. Существуют разные определения больших данных, но большинство из них базируется на концепции «пяти V» больших данных:

- **Volume** (рус. объем) – оперирование информацией, объем которой начинает измеряться от гигабайт и не может быть обработан на одном устройстве.

- **Variety** (рус. разнообразие) – обрабатываемая информация имеет самую разнообразную природу, это как структурированная, так и неструктурированная информация.

- **Velocity** (рус. скорость/оперативность) – скорость генерирования данных достаточно высока для оперативного анализа в режиме real-time.

- **Veracity** (рус. достоверность) – для «больших данных» характерна достоверность, а также возможность идентификации объекта мониторинга.

- **Value** (рус. ценность) – аналитические выводы имеют непосредственную ценность для организаций (государственных и бизнес-структур), так как предоставляют ценные инсайты об объектах мониторинга

(англ. insights – интуиция). «Большие данные» имеют высокий потенциал монетизации.

Большие данные тесно связаны с искусственным интеллектом в узком понимании (от англ. Artificial Intelligence), статистическим моделированием (Statistical Modeling) и наукой о данных (англ. Data Science), а также машинным обучением (англ. Machine Learning) и глубоким обучением на основе искусственных нейронных сетей (от англ. Deep Learning). Концептуально большие данные опираются на закон больших чисел, согласно которому, чем выше объем выборочной совокупности, тем лучше она описывает генеральную совокупность. Большие данные позволяют моделировать поведение системы, в которой присутствуют сложные многоуровневые отношения между ее структурными элементами. Основное, что необходимо для функционирования систем обработки больших данных – это развитая инфраструктура данных, а также постоянная актуализация данных. Именно поэтому первыми, кто активно начал развивать сервисы Big Data стали крупные технологические компании с большими аудиториями, например, Google Cloud, Amazon Web Services, Facebook Analytics и другие.

В связи с тем, что из-за значительного объема, данные невозможно обрабатывать на единичных узлах вычислительной инфраструктуры, обработка происходит на базе облачной инфраструктуры, то есть на основе компьютерных кластеров. Ключевым принципом работы с большими данными является технология MapReduce, которая состоит из 3-х основных этапов: на этапе Map происходит предобработка данных и распределение подзадач на кластерную инфраструктуру, на втором этапе данные обрабатываются параллельно, на третьем этапе Reduce происходит свертка промежуточных результатов в финальный результат.

Мощный толчок развитию больших данных придает динамичное наращивание доступных компьютерных мощностей, развитие облачной сетевой инфраструктуры и технологий параллельной обработки данных, а также, что существенно, инструментария моделирования. В связи с тем, что телекоммуникационные компании, как правило, являются провайдерами облачных решений, обладают значительными вычислительными мощностями для поддержания инфраструктуры, а также опираются на данные в реализации бизнес-стратегий, компании индустрии находятся в авангарде имплементации технологий обработки больших данных.

Используя свои собственные данные и комбинируя данные из разных источников, телеком-компании стремятся к более глубокому пониманию взаимодействия своих клиентов, более эффективному управлению жизненным циклом продуктов, а также построению детализированной картины бизнеса в целях улучшения качества обслуживания клиентов и извлечения ценности из имеющихся данных.

1.2.2 Машинное обучение и искусственный интеллект

Ключевыми драйверами перехода к менеджменту на основе данных являются методы машинного обучения и развитие нейросетевых подходов моделирования, позволяющих извлекать ценные для бизнеса выводы из данных различного представления, таких как данные в табличном виде, текстовые данные, изображения и аудио-данные.

Искусственный интеллект (англ. Artificial Intelligence) занимает важное место в процессах интеллектуальной обработки данных, так как построенные на его основе модели, применяются для автоматизации систем поддержки принятия решений. Однако, в бизнес-контексте телекоммуникаций, несмотря на общеупотребимость, некорректным является использование термина “искусственный интеллект”, так как он носит окраску “в широком понимании”, что не соотносится с подходами, применяемыми в индустрии. Вместо этого более корректным является употребление термина “машинное обучение”, которое означает комплекс статистических техник и интеллектуальных алгоритмов обработки данных, используемых для моделирования различных процессов, и является подразделом искусственного интеллекта. Термины, связанные с областью искусственного интеллекта, такие как машинное обучение, глубокое обучение, data-майнинг часто используются взаимозаменяемо, что является ошибочным. Такая дисциплина информатики, как машинное обучение – средство достижения искусственного интеллекта, в то время как глубокое обучение – один из подразделов машинного обучения, под которым, как правило, понимается применение глубоких нейронных сетей для обработки данных со сложными нелинейными зависимостями. Хорошим примером задач, решаемых посредством методов глубокого обучения являются: компьютерное зрение, системы машинного перевода, распознавание образов, поиск закономерностей в данных, распознавание речи, обработка человеко-читаемого текста и др.

При решении задач моделирования и прогнозирования с помощью машинного обучения, их формализуют при помощи двух концептуальных подходов: обучение с учителем (англ. supervised) и обучение без учителя (англ. unsupervised). Различие двух подходов заключается в том, что для supervised подходов используются размеченные данные, решаются задачи классификации и регрессии. В unsupervised подходах применяются неразмеченные данные для кластеризации и понижения размерности многомерных данных [17].

Инструментарий машинного обучения доказал свою эффективность в освоении телекоммуникационными компаниями потенциала накопленных данных: планирование развертывания телекоммуникационной сети, управление клиентским опытом, сопровождение клиентов, управление жизненным циклом продуктов, построение поведенческих моделей, проведение маркетинговых кампаний, операционный менеджмент и другие.

Таким образом, машинное обучение является неотъемлемой частью бизнес-моделей, основанных на данных, реализуемых телекоммуникационными компаниями.

1.2.3 Интернет вещей

Интернет объединяет не только людей, но и оборудование. Кибернетические устройства глубже проникают в повседневную жизнь, создают данные, регистрируют события. В 2009 году, по информации компании Cisco, количество подключенных устройств превысило количество людей в интернете, это стало моментом перехода от «интернета людей» к «интернету вещей». На данный момент также говорят о создании ценности подключенными к интернету вещами, поэтому также формируется концепция «интернета ценности», в основе которой лежат технологии распределенного реестра. В обзоре PwC отмечается, что развитие интернета вещей было бы невозможно без развития комплекса технологий Big Data [18]. Развитие интернета вещей находится в тесном сопряжении со сферой телекоммуникаций и во многом ей определяется.

Интернет вещей – новый этап развития сети интернет, который значительно расширяет возможности сбора, анализа и обработки данных, без включения человека в эту цепочку. Человек же использует информацию, переработанную IoT-системами из потока данных, для принятия решений. Несмотря на то, что концептуально интернет вещей оформился только в 2008 году, активно внедряются соответствующие подходы в самые различные сферы: сенсорные системы, средства связи, программные платформы для организации промышленных комплексов, умных сетей электропередач, платформы аналитических решений, объекты «умной» городской инфраструктуры, системы хранения и обработки данных. Потенциал расширения коннективности между людьми посредством средств связи ввиду насыщения рынка можно считать освоенным, в то время как неосвоенным остается ценность, возникающая в результате объединения вещей в интеллектуальные сети.

Как и для любой другой технологии, для IoT актуальны проблемы, связанные с технической реализацией и социо-экономическими эффектами внедрения. Основная угроза исходит из понимания того, что ранее, те приборы, которые сейчас становятся элементами IoT-сетей, не имели информационной составляющей. Важным условием внедрения IoT, как и для любой другой технологии, является наличие законодательства, одновременно ориентированного на раскрытие технологического потенциала и на защиту интересов граждан и потребителей.

Многие телекоммуникационные компании приступили к реализации проектов, связанных с интернетом вещей. Естественные конкурентные преимущества провайдеров телекоммуникационных решений и накопленный опыт реализации масштабных инфраструктурных проектов оптимально

встраивают интернет вещей в продуктовое портфолио компаний рынка телекоммуникаций. М2М-решения (от англ. machine to machine) представляют собой быстрорастущий сегмент телекоммуникационного рынка, диверсифицирующий источники дохода компаний.

Создание стоимости в форме продуктов и услуг породило понятие «цепочки создания ценности» - это последовательность действий, посредством которых организация преобразует входные ресурсы в экономические результаты. Аналогичным образом, реализация полного потенциала IoT обуславливает необходимость оценки источников генерации ценности из информации. Представленная ниже схема, разработанная компанией Deloitte, описывает процесс создания ценности интернетом вещей [19]:

Схема имеет циклический вид, так как аналитическая ценность обрабатываемых данных, инициирующая дальнейшие действия, является формой обратной связи между реальным миром и информационными системами, в основе которых лежат отраслевые стандарты IoT. Проходя различные фазы цикла создания ценности, информация обрабатывается посредством различных технологических подходов. В первую очередь, событие в IoT-сети регистрируется сенсорной системой, создающей формализованную информацию о событии. Далее эта информация коммуницируется и агрегируется, тем самым производится аналитический продукт. Цикл создания ценности завершается созданием отклика автономной сети или аналитический результат напрямую участвует в процессе принятия решений человеком. Таким образом, результирующим событием цикла, является адаптация поведения IoT-сети как на уровне отдельных компонентов, так и в целом.

Анализ абстрактной схемы цепочки создания ценности из информации, генерируемой в сети интернета вещей, позволяет сделать вывод о высоком потенциале разработки, интеграции и внедрения IoT-решений телекоммуникационными компаниями для различных сфер применения, так как сопутствующая изменчивость в поведении систем позволяет оптимизировать структуру издержек компании, повысить оперативность и гибкость принимаемых решений, а также автоматизировать рутинные операции администрирования устройств.

1.2.4 Роботизированная автоматизация процессов

Роботизированная автоматизация процессов (от англ. Robotic Process Automation или RPA) - ряд технологических подходов, применяемых для автоматизации рутинных бизнес-процессов в целях оптимизации временных и финансовых затрат. RPA-системы подразумевают наличие специализированной платформы для построения алгоритмов действий, на базе которой разработчик дает виртуальному агенту (роботу) четкие инструкции. Также в последнее время функционал RPA-платформ

расширяется инструментарием машинного обучения. Это позволяет применять нечеткую логику при решении роботом различных задач. Таким образом, происходит автоматизация рутинного труда, совершаемого человеком, путем настройки или обучения роботизированных автоматизационных систем. RPA-системы также функционируют на основании эмуляции взаимодействия пользователя с графическим интерфейсом информационной системы. Автоматизация может включать в себя обработку данных и взаимодействие между различными цифровыми системами.

RPA-системы выступают важнейшим фактором снижения операционных издержек бизнеса, а также являются залогом быстрой адаптации бизнеса к новым условиям функционирования.

По результатам исследования, проведенного компанией Ernst & Young телеком-компаниями сталкиваются со следующими вызовами [9]:

- Низкий уровень оперативности бизнеса (необходимость в настройке различных информационных систем, обеспечивающих операционную деятельность);
- Неэффективность потока информации (необходимость в синхронизации и координации работы различных подразделений);
- Высокие операционные издержки и капиталозатраты (необходимость в управлении издержками компании);

Применение RPA в целях оптимизации бизнес-процессов в сфере телекоммуникаций целесообразно в связи с возможностью горизонтального масштабирования автоматизируемых процессов на заинтересованные подразделения компании.

В сфере телекоммуникаций существует множество направлений, в которых ярко выражена необходимость выполнения рутинных и репитативных операций по определенным алгоритмам. В качестве примера можно привести следующие кейсы: службы клиентской поддержки, операционная деятельность колл-центров, управление и планирование развертывания сети, внесение данных в CRM-систему, актуализация данных, построение регулярной отчетности и многие другие задачи. RPA также косвенно влияет на качество менеджмента клиентского опыта, так как в случае автоматизации рутинных задач, представитель компании сможет уделять больше внимания взаимодействию с клиентом, нежели администрированию информационных систем. Системы роботизированной автоматизации иногда рассматриваются в качестве альтернативных используемым КИС: так как внедрение корпоративной системы, как правило, является чрезвычайно затратным и полный функционал этой системы часто остается невостребованным, с RPA появляется возможность использования более простых систем, усложняя их средствами RPA в случае необходимости. Среди преимуществ RPA-систем можно выделить следующие: повышение скорости выполнения работ, быстрый возврат

инвестиций, высокие возможности масштабирования, высокая точность работы, качество и безопасность.

Наиболее популярными являются следующие RPA-платформы: Automation Anywhere, Blue Prism, electroNeek, Kofax, Kryon, ROBIN и Softomotive [20].

Таким образом, вариативность сценариев и бизнес-контекстов в сфере телекоммуникаций, в которых могут быть задействованы RPA-решения, делает этот набор подходов и практик крайне ценным для бизнеса. RPA оценивается аналитиками, как одно из ключевых технологических решений, формирующих будущее рынка телекоммуникаций [21].

1.3 Инновационные конкурентные стратегии в индустрии телекоммуникаций

Телекоммуникационный бизнес становится более динамичным и высокотехнологичным, компании внедряют новые продукты, а также проводят реинжиниринг бизнес-процессов. На основании проведенного выше анализа можно сделать выводы, касательно факторов обуславливающих трансформацию бизнес-моделей телекоммуникационных компаний:

- внедрение инновационных технологических подходов (аналитика больших данных, машинное обучение, интернет вещей и роботизированная автоматизация процессов) и повышение технологической зрелости компаний;
- изменения конъюнктуры мирового рынка телекоммуникаций, характерные для фазы перехода от третьей к четвертой фазе развития рынка;
- кросс-индустриальная специфика инноваций, определяющая высвобождение ценности “на стыке” индустрий;
- изменение потребностей рынка, связанных с мега-трендами;
- масштабная цифровизация экономики;

Для понимания перспектив развития рынка, анализ факторов трансформации является необходимым, но недостаточным условием. Также важным является вопрос того, как в контексте существующих бизнес-моделей, телекоммуникационные компании интегрируют новые подходы. Адаптация компаний к новым условиям принимает форму имплементации инновационных бизнес-моделей, которые будут рассмотрены ниже.

Компании диверсифицируют бизнес путем проведения вертикальной интеграции, придерживаются контентной стратегии, выпускают партнерские продукты с кредитно-финансовыми организациями, формируют новые предложения, опираясь на накопленные данные, внедряют цифровые каналы взаимодействия с клиентами, а также используют конвергентные стратегии.

Рассмотрим некоторые популярные стратегии развития телекоммуникационных компаний.

1.3.1 Контентная стратегия

Наблюдается бурный рост рынка ОТТ-сервисов, что отражается в нарастающем разрыве между доходностью индустрии телекоммуникаций и ОТТ. Фактически происходит миграция основной отраслевой ценности от телекоммуникационных операторов к ОТТ-провайдерам, которые предоставляют услуги связи клиентам и другие цифровые сервисы (приложения, голосовые звонки и мессенджеры) на основании ресурсов телеком-операторов. В течение последнего десятилетия телекоммуникационные компании испытывали снижение доходов, денежного потока, а также окупаемости инвестиций, в то время как технологические компании, такие как Google, Facebook, Amazon и другие, демонстрировали стремительный рост, закладывая в основу собственного бизнеса инфраструктуру телеком-операторов. Само наличие этой устоявшейся и функционирующей инфраструктуры перестает восприниматься как инновация, уступая место сервисам, которые используют ресурсы этой сети.

Для управления источниками дохода телекоммуникационных компаний актуальным является сдвиг в структуре потребления трафика. Происходит переход от роста за счет предоставления голосового мобильного трафика (третья фаза) к интернет-трафику, в частности к видео-трафику. Игровая индустрия и активно развивающийся рынок платформ предоставления контента также оказывают значительное влияние на развитие телекоммуникационной индустрии.

Также устаревает консервативная формула лидерства на рынке телекоммуникаций, согласно которой владение сетевой инфраструктурой и наличие обширной дистрибьюторской сети являлись основными конкурентными преимуществами. На данный момент они начинают играть меньшую роль в продуктовой дифференциации. ОТТ-сервисы имеют повышенную ценность для конечного потребителя, так как являются ключевыми как с точки зрения генерации трафика в сети телеком-оператора, так и создания добавленной стоимости, а также освоения ценности от сетевого эффекта. Фактически, наблюдается ужесточение конкуренции между телеком-компаниями и ОТТ-провайдерами.

Телекоммуникационные компании стремятся получить доход от предложения новых сервисов, поэтому, чтобы компенсировать коммодитизацию своих традиционных услуг связи, они часто прибегают к контентной стратегии. Она подразумевает предоставление видео- и аудио-контента клиентам через платформы, в том числе и контента собственного производства.

Телекоммуникационные компании могут быть как конкурентами, так и партнерами ОТТ-бизнеса в зависимости от выбранной стратегии. Немногие потребители хотят пользоваться множеством ОТТ для получения

интересного им контента, поэтому агрегация контента - это востребованная услуга, которую могут предоставлять телекоммуникационные компании. Это дает возможность эффективно конкурировать с крупными провайдерами контента, такими, как Netflix или Amazon. Также телекоммуникационные компании могут установить партнерские отношения с этими агрегаторами для доступа к большим объемам контента.

Чтобы эффективно использовать эту возможность, существует три основных, но не обязательно взаимоисключающих подхода к предоставлению видеоконтента:

- создание контента самостоятельно через собственные производственные отделы;
- приобретение контента напрямую у производителей контента;
- предложение контента агрегаторов, таких как Netflix и Amazon.

Однако чаще всего телеком-операторы используют смесь всех трех подходов для достижения максимального эффекта.

1.3.2 Предоставление data-продуктов

Телеком операторы являются не только поставщиками собственных цифровых продуктов и услуг, но и представляют собой ценный фактор содействия в диджитализации для других секторов.

Данные, генерируемые телекоммуникационными компаниями, имеют большую ценность также и для других прикладных сфер, таких, как банковская сфера, здравоохранение, градостроение, образование и планирование городской инфраструктуры и др. В связи с этим телекоммуникационные компании начинают предоставлять так называемые data-продукты для других индустрий, то есть сервисы, в основе которых лежат данные, накапливаемые о поведении клиентов телекоммуникационных компаний и обладающие высокой ценностью для решения бизнес-задач, возникающих в других индустриях. Например, для задач планирования, развития и оптимизации городской среды высокую ценность имеют геолокационные данные о местах высокой концентрации людей, а также об их передвижении. Банковская сфера сталкивается с потребностью в надежных данных для управления и прогнозирования рисков при работе с заемщиками, поэтому телекоммуникационные компании начинают предлагать скоринговые модели для банковских продуктов на основании телеком-предикторов, повышающие прогностическую силу банковских скоринговых моделей. Помимо этого различные индустрии, такие как банковская сфера, страхование и общественная безопасность берут на вооружение множество подходов для борьбы с мошенничеством и его оперативным выявлением. Для решения подобных задач велика ценность различных типов данных, которые генерируются телекоммуникационными компаниями. Количество контекстов и сфер применения телекоммуникационных данных стремительно растет, в связи с чем

возрастает потребность в эффективных системах администрирования данных.

1.3.3 Финансовые продукты

В последнее время наблюдается процесс конвергенции банковской сферы и телеком-операторов. Деактуализация консервативных бизнес-моделей в обеих сферах подталкивает бизнес на поиск новых источников дохода и роста. Помимо существенных конъюнктурных изменений, происшедших в индустриях на протяжении последних 5 лет, также стоит принять во внимание развитие смежных рынков и становление новых технологических подходов. В качестве определяющих для процессов подобной конвергенции факторов следуют выделить следующие:

1. Рост потребностей телекоммуникационных операторов в финансовых ресурсах в связи с необходимостью развертывания сетей нового поколения (IoT - интернет вещей, 5G - стандарт высокоскоростной мобильной передачи данных);
2. Технологизация и автоматизация бизнес-процессов в банковской и страховой сферах (FinTech - финансовые технологии, InsureTech - страховые технологии), а также развитие цифровых каналов взаимодействия с клиентами;
3. Развитие экосистемы данных и углубленной аналитике как в банкинге, так и в телеком, позволяющей выявлять потребности клиентов и формировать персонализированные предложения;
4. Становление благоприятных регуляторных условий;
5. Потребность обеих индустрий в качественной трансформации бизнес-моделей;
6. Тренд в телеком-индустрии в направлении формирования конвергентных предложений (расширение продуктового микса новыми ценностными предложениями);

На данный момент обе индустрии сталкиваются со схожими задачами, к решению которых может привести консолидация ресурсов и создание схем совместной коммерциализации. Как показывает, интеграция финансовых сервисов в портфель услуг, предоставляемых телеком-операторами, показывает сдержанные темпы роста ввиду наличия барьеров входа на рынок банкинга. Среди них можно выделить высокие регуляторные требования, отсутствие развитой системы страхования вкладов, неразвитость каналов финансирования и необходимость обеспечения высокой степени безопасности [22].

Мобильные операторы предлагают широкий спектр посреднических и финансовых сервисов для клиентов. Среди них можно выделить такие услуги, как программы рассрочки и отложенных платежей, микрокредитование, микрострахование, а также мобильный банкинг, интегрирующий различные сервисы.

В связи с насыщением рынка смартфонов, распространением мобильных платежных сервисов и распространением девайсов, поддерживающих технологию NFC, получили широкое распространение мобильные платежи. Телекоммуникационные компании начинают предлагать финансовые услуги, однако, пока не готовы предоставлять полноценные финансовые сервисы, включающие сберегательные и кредитные продукты.

Можно выделить следующие вызовы, характерные для обеих индустрий, представляющих собой точки сближения для телеком-компаний и банков:

- Снижение маржинальности бизнеса, обусловленное эволюцией рынка;
- Рост конкуренции;
- Необходимость в глубоком реинжиниринге бизнес-процессов в целях соответствия потребностям клиентов;

Для выхода на рынок банкинга у телекоммуникационного оператора есть несколько опций: в первую очередь следует выделить партнерство с функционирующими кредитно-финансовыми учреждениями, кроме того для телеком-операторов доступна возможность создания банка, однако существует значительный порог входа на рынок.

Рассмотрим различные способы реализации конвергентной стратегии с позиции банков. Для выхода банков на рынок телекоммуникаций, им доступны следующие стратегии развития: использование инфраструктуры существующих провайдеров, предоставление услуг мобильной связи на базе виртуального оператора и создание собственной инфраструктуры. Как правило, использование инфраструктуры провайдеров редко реализуется на практике в связи с возникающим высоким уровнем зависимости от партнера-провайдера, создание собственной инфраструктуры также редко реализуется из-за высоких барьеров входа на рынок, поэтому, запуск услуг связи на базе виртуальных операторов мобильной связи (MVNO - mobile virtual network operator) является наиболее оптимальным вариантом с точки зрения предоставления банковскими структурами услуг мобильной связи и внедрением продуктовых решений. Это можно объяснить тем, что мобильный оператор берет на себя обязательства по развертыванию необходимой инфраструктуры, сокращая издержки и время на создание нового сервиса. MVNO или виртуальный оператор - схема предоставления услуг мобильной связи, в которой компания не строит собственную сеть, а арендует ее у базовых операторов, что фактически представляет собой оптовую закупку трафика у владельца сети с возможностью перепродавать его под собственным брендом [23].

Таким образом, подобные совместные проекты имеют большой потенциал с точки зрения совершенствования аналитических процессов конвергентной организации. Ценность подобного сотрудничества между банками и телеком-операторами на различных временных горизонтах представлена в таблице 1.1.

Таблица 1.1 – Преимущества от реализации конвергентной стратегии банков и телекоммуникационных компаний

	Банки	Telco
Краткосрочный (реализация совместных проектов и партнерских программ)	<ul style="list-style-type: none"> • Повышение непроцентного дохода; • Снижение издержек на сопровождение и удержание клиентской базы; 	<ul style="list-style-type: none"> • Арендный доход от эксплуатации банками сетевой инфраструктуры; • Предоставление новых финансовых услуг;
Среднесрочный (кросс-предложение продуктов и интеграция клиентской базы)	<ul style="list-style-type: none"> • Усиление позиции банков в сегменте мобильных платежей и розничном сегменте в целом; • Реализация цифровой мобильной стратегии; • Рост клиентской базы за счет кросс-предложения услуг; • Увеличение проникновения; 	<ul style="list-style-type: none"> • Повышение лояльности клиентов; • Совершенствование менеджмента абонентской базы; • Совершенствование механизма принятия бизнес-решений; • Повышение финансовой устойчивости;
Долгосрочный (полная диагональная интеграция)	<ul style="list-style-type: none"> • Обогащение корпоративных источников данных для углубленной аналитики; • Имплементация технологии обработки больших данных; • Формирование единой экосистемы, включающей разнообразный продуктовый микс. 	

Примечание – Разработка автора

Объединение корпоративных информационных систем различных классов (CRM, ERP, BPM и других) и слияние клиентских баз, а также построение интегрированных аналитических систем - несомненно, весомые конкурентные преимущества. Размывание границ между банками и мобильными операторами существенно трансформирует эти рынки и становится единым этапом эволюционного развития обеих индустрий. На данный момент рано говорить о том, какая из индустрий демонстрирует большую мотивацию участия в конвергентном процессе, тем не менее наличие успешных реализаций партнерских программ со стороны банков, говорит о том, что именно банковская сфера является более заинтересованной в конвергенции с мобильными операторами и имеет для этого более благоприятные условия.

Выводы по главе 1

Проанализировав различные технологические подходы с точки зрения их значимости в контексте текущих тенденций развития индустрии телекоммуникаций, можно сделать следующие выводы и дать рекомендации, касательно их внедрения телекоммуникационными компаниями:

Во-первых, следует отметить изменение глобальной конъюнктуры телекоммуникационного рынка. Современные телекоммуникационные компании используют новые конкурентные стратегии, основанные на данных. Также необходимо отметить, что формула лидерства, практикуемая телекоммуникационными компаниями на третьей фазе развития рынка и выражающаяся в контроле инфраструктуры, неизменно устаревает на четвертой фазе. Качество услуг и клиентский опыт, предоставляемый компаниями, являются важнейшими дифференциальными факторами на рынке. Наиболее значимыми современными трендами развития телекоммуникационной индустрии являются: диверсификация проектного портфеля, поиск новых источников выручки, а также цифровая трансформация бизнеса.

Во-вторых, в телекоммуникационной индустрии под влиянием интенсивного внедрения таких технологических подходов, как: аналитика больших данных, машинное обучение и искусственный интеллект, интернет вещей, а также роботизированная автоматизация процессов, происходит трансформация цепочек создания отраслевой ценности. Если ранее единственным источником ценности, создаваемой телеком-компаниями, были предоставляемые ими услуги, то постепенно произошло добавление нового источника ценности, основанном на многообразии данных и выводах, полученных из данных при помощи технологий анализа больших объемов данных.

В-третьих, телекоммуникационные услуги подвергаются коммодизации, что обуславливает необходимость внедрения инновационных бизнес-стратегий, способных коренным образом трансформировать восприятие конечного потребителя телеком-компаний. В связи с этим компании на рынке телекоммуникаций предпринимают усилия по превращению из провайдеров услуг в провайдеров конвергентных сервисов, интегрированных в единую экосистему. Примером таких экосистемных решений являются три рассмотренных бизнес-стратегии: контентная стратегия, предусматривающая создания медиа-платформы, предоставление data-продуктов, а также стратегия предоставления услуг финансового посредничества.

Все эти аспекты функционирования рынка телекоммуникаций актуализируют задачи управления данными как активом, а также подчеркивают важность поиска путей повышения эффективности процессов управления данными и реализации их потенциала.

ГЛАВА 2 АНАЛИТИЧЕСКИЕ ЭКОСИСТЕМЫ ТЕЛЕКОММУНИКАЦИОННЫХ КОМПАНИЙ

2.1 Анализ методологических подходов управления данными в телекоммуникационных компаниях

Телекоммуникационные компании предпринимают множество усилий для повышения эффективности управления данными. На четвертом этапе развития рынка телекоммуникаций данные являются не только фактором дифференциации, но и имеют высокий потенциал к коммерциализации. Для извлечения ценности из данных необходимо наличие аналитической инфраструктуры, которая позволяет в оперативном режиме трансформировать сырые данные в ценный актив компании.

Многие организации признают, что их данные являются жизненно важным активом, определяющим конкурентное положение компании и эффективность ее бизнес-процессов. Данные и информация дают им представление о клиентах и предоставляемых продуктах и услугах. Глубокое понимание специфики бизнеса на основании аналитики данных способствует внедрению инноваций и достижению стратегических целей. Несмотря на это, лишь немногие организации эффективно управляют данными как активом. Извлечение ценности из данных невозможно без тщательного структурирования и планирования процессов работы с данными. Также необходимо понимать место аналитических систем в контексте целостного подхода управления данными, объединяющего все аспекты цепочки извлечения ценности из данных.

Управление данными (от англ. Data Management) – это мероприятия, направленные на разработку, исполнение и контроль планов, политик, программ и практик, которые обеспечивают, контролируют, защищают и повышают ценность данных и информационных активов на протяжении их жизненных циклов.

Активности, предусмотренные концепцией управления данными разнообразны и включают как разработку стратегических подходов извлечения ценности из данных, так и операционные вопросы управления данными.

Выделяют несколько типов подходов, направленных на управление данными компании как активом:

1. Методологии управления данными;
2. Методологии анализа данных;
3. Организационные структуры.

С целью выделения лучших практик для внедрения в телекоммуникационные компании данные подходы будут рассмотрены более детально.

2.1.1 Методологии управления данными

Существует множество методологических подходов к управлению данными, однако на практике преимущественно используются 2 из них: методология управления данными, предложенная международной ассоциацией управления данными DAMA DMBOK, и методология оценки степени зрелости инфраструктуры управления данными от института зрелости интеграции (Capability Maturity Model Integration Institute) CMMI DMM.

Для оценки потенциала применения данных методологий, проведем сравнительный анализ их ценности для сферы телекоммуникаций.

DAMA DMBOK

Методология впервые появилась в 2009 году с публикацией организацией DAMA свода знаний, посвященных управлению данными – DMBOK (от англ. Data Management Body Of Knowledge). Данная методология прошла проверку временем и в 2017 году была выпущена актуализированная версия DMBOK 2, учитывающая современный опыт построения экосистем управления данными. Суть методологии DMBOK наиболее емко можно представить в виде секторной диаграммы, состоящей из 11 сегментов, представляющих целостную систему взаимосвязанных компонентов управления данными [24].

Диаграмма представлена ниже (Рисунок 2.1).

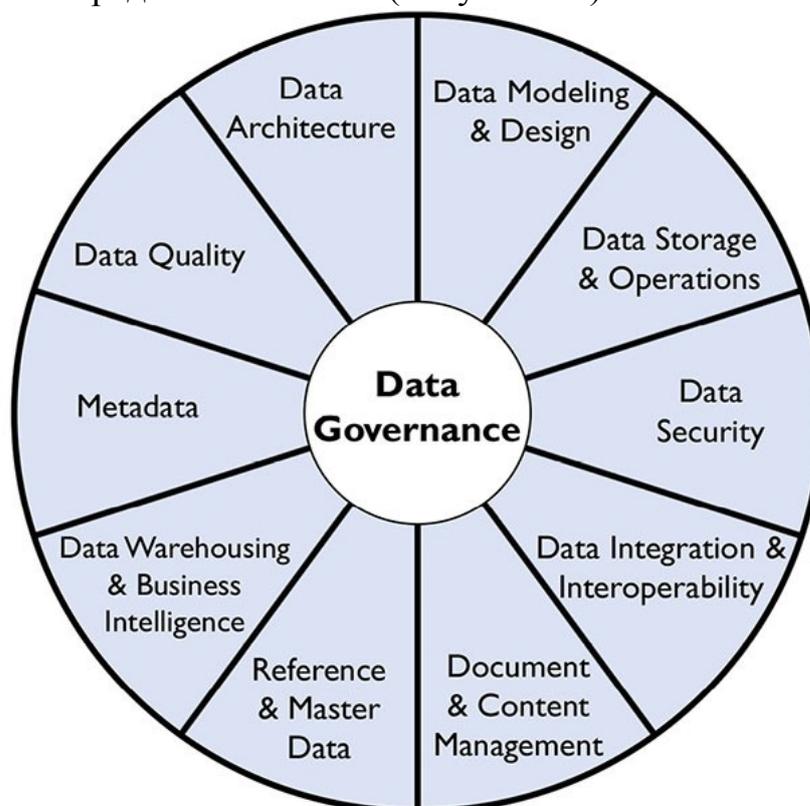


Рисунок 2.1 – Секторная диаграмма компонентов методологии DAMA-DMBOK
Примечание – Источник: DAMA-DMBOK (<https://dama.org/content/body-knowledge>)

В методологии DMBOOK компонент администрирования данных (от англ. Data Governance) занимает центральное место, так как на базе него выстраиваются все остальные компоненты, которые подчинены практикам администрирования данных в компании. Целью компонента администрирования данных является обеспечение соблюдения всех практик управления данными, выделенных компанией в качестве стратегических. В то время как методология DMBOOK в целом ориентирована на определение способов извлечения ценности из данных компанией, компонент Data Governance сфокусирован на решениях, принимаемых касательно управления данными, а также на настройку бизнес-процессов, связанных с управлением данными.

Остальные 10 компонентов DMBOOK представляют собой взаимосвязанные самостоятельные области, каждая из которых релевантна для построения эффективной экосистемы управления данными.

Рассмотрим отдельные компоненты методологии DMBOOK более детально:

- Архитектура данных (от англ. Data Architecture) – компонент, отвечающий за определение необходимых уровней абстракции бизнес-процессов компании; Архитектура данных определяет набор требований к системам управления данными, а также стратегию управления данными как активом;
- Моделирование данных (от англ. Data Modeling and Design) – компонент, ориентированный на изучение и анализ данных, а также извлечение ценности из данных;
- Хранение данных (от англ. Data Storage and Operations) – компонент, отвечающий за дизайн, внедрение и поддержку хранимых данных с целью максимизации ценности данных на всем их жизненном цикле от создания/приобретения до удаления;
- Безопасность данных (от англ. Data Security) – компонент, отвечающий за разработку, внедрение и имплементацию политик безопасности данных, а также защиту информационных активов компании;
- Интеграция данных (от англ. Data Integration and Interoperability) компонент, описывающий процессы, связанные с передачей данных, использованием данных различными отделами компании, а также консолидацией данных для различных бизнес-задач;
- Документооборот и контентная стратегия (от англ. Document and Content Strategy) – компонент методологии, отвечающий за контроль и управление данными компании, хранимыми вне реляционных баз данных, таких как документация, а также неструктурированные данные;
- Мастер-данные (от англ. Master Data) – компонент DMBOOK, который отвечает за высокоуровневый слой абстракции управления данными, а

также предоставляет информацию о более низкоуровневых уровнях абстракции данных компании;

- Бизнес-анализ (от англ. Business Intelligence) – компонент, отвечающий за структурирование и логику данных, а также эффективное представление данных и построение моделей данных;
- Метаданные (от англ. Metadata) – компонент ДМВОК, который описывает структурированные данные, представляющие собой характеристики описываемых сущностей для целей их идентификации, поиска, оценки и управления ими;
- Качество данных (от англ. Data Quality) – компонент методологии ДМВОК, отвечающий за качество используемых данных, а также процессы управления качеством;

Помимо отдельных компонентов ДМВОК также позволяет оценить роль человека в процессах управления данными, что дает возможность трансформировать эту методологию в элемент корпоративной культуры, таким образом подкрепив процесс принятия решений в компании когерентной операционной моделью [24].

СММІ DMMM

В 2014 году институт СММІ выпустил методологию построения модели зрелости управления данными (Data Management Maturity Model). Эта модель позволяет оценить степень зрелости развития экосистемы управления данными по пяти ключевым параметрам. Ключевой особенностью методологии СММІ DMMM является применение концепции уровня зрелости к различным измерениям экосистемы управления данными компании, что позволяет оценить текущий уровень зрелости и определить траекторию движения к более высоким уровням зрелости [25].

Сущностная характеристика методологии СММІ DMMM может быть представлена в виде диаграммы:

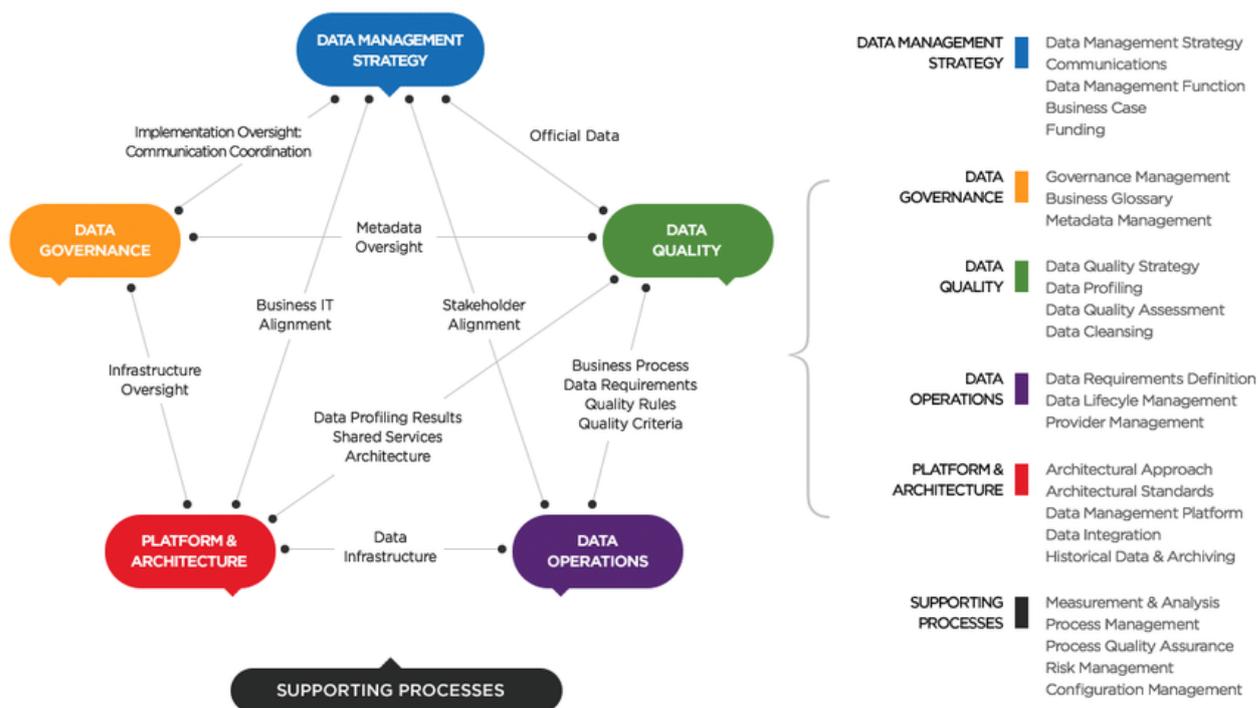


Рисунок 2.2 – Диаграмма компонентов методологии CMMI DMMM

Примечание – Источник: CMMI Institute (<https://cmminstitute.com/data-management-maturity>)

В CMMI DMMM основными единицами анализа являются компоненты управления данными:

- **Стратегия управления данными** (от англ. Data Management Strategy) представляет собой анализ подходов к оптимизации и совершенствованию стратегической программы управления данными компании;
- **Администрирование данных** (от англ. Data Governance) является набором практик, определяющих способы построения высокоэффективной экосистемы управления данными, ориентированной на масштабное вовлечение всех подразделений бизнеса в цифровизационные процессы и имплементацию data-ориентированных подходов в бизнес-процессы;
- **Качество данных** (от англ. Data Quality) представляет собой компонент методологии, отвечающий за оценку качества данных, способы контроля качества и валидности данных, а также механизмы повышения качества данных;
- **Операционное управление данными** (от англ. Data Operations) – компонент CMMI DMMM, ориентированный на обеспечение прозрачности процессов управления данными, а также обеспечивающий соблюдение требований на всех этапах жизненного цикла данных;

- **Платформа и архитектура** (от англ. Platform and Architecture) – компонент методологии, позволяющий определить оптимальную технологическую платформу для обработки данных.

Сравнительный анализ методологий управления данными

Для определения методологии, наиболее подходящей для внедрения в телекоммуникационную сферу, необходимо провести сравнительный анализ.

Преимущества методологии DAMA DMBOK:

- В первую очередь, подход, предложенный DAMA ориентирован на интерпретацию данных как экономического актива компании, что позволяет непосредственно проецировать конкретные мероприятия по развитию инфраструктуры управления данными на экономические результаты работы компании и ее эффективность;
- Концепция DAMA позволяет дать целостную оценку работы различных компонентов системы управления данными, определить взаимосвязи между ними, а также выявить точки роста эффективности управления данными;
- Универсальность концепции дает возможность описать бизнес-процессы компаний, которые в различной степени интегрируют анализ данных в свои бизнес-модели;
- Концепция позволяет оценить неэффективные бизнес-процессы и дать практические рекомендации для их совершенствования.

Недостатки DAMA DMBOK:

- Сложность методологии не позволяет быстро получить практический результат;
- Как правило, практические рекомендации на основании DMBOK необходимо закладывать в бизнес-процессы с начала развития компании.

Преимущества методологии CMMI DMMM:

- CMMI DMMM имеет достаточно простую структуру;
- Методология позволяет определить текущий статус зрелости экосистемы управления данными;
- В методологии заложено множество предписательных элементов, позволяющих определить необходимый набор действия для оптимизации бизнес-процессов;
- CMMI DMMM позволяет произвести быструю предварительную оценку процессов управления данными.

Недостатки CMMI DMMM:

- Как правило, телекоммуникационный бизнес имеет довольно развитую экосистему управления данными, поэтому компонентов CMMI DMMM недостаточно для детализации траектории развития;
- В отличие от других методологий управления данными, CMMI DMMM не рассматриваются несколько существенных областей, таких как углубленная аналитика данных и управление мастер-данными, что является существенным недостатком с точки зрения применимости этой концепции в современных телекоммуникационных компаниях, так как аналитика данных является ключевым драйвером роста телеком-рынка и требует более детального подхода.

Таким образом, с учетом рассмотренных выше преимуществ и недостатков методологий, более оптимальной для внедрения телекоммуникационными компаниями на четвертом этапе развития рынка является методология DAMA DMBOK. Данный подход масштабируем и является гибким по отношению к различным инновационным бизнес-стратегиям телеком-компаний, так как оперирует обособленными базовыми элементами экосистемы управления данными.

В контексте рассмотренной целостной системы управления данными на основании методологии DMBOK необходимо оценить характер связей аналитической экосистемы с отдельными компонентами методологии и определить наиболее важные компоненты, непосредственно влияющие на построение аналитических процессов в телекоммуникационных компаниях.

Среди 10 компонентов DMBOK можно выделить 3 из них, оказывающие наибольшее влияние на эффективность аналитических процессов в компании. Важнейшим компонентом является качество данных, так как именно он обуславливает валидность сделанных на их основе выводов, качество статистических моделей и аналитическую ценность. Также необходимо отметить, что BI-системы неразрывно связаны с аналитикой, так как именно на их основе происходит структурирование данных и трансформация сырых данных в аналитические агрегаты, пригодные для использования в приложениях углубленной аналитики. Также необходимо выделить интеграцию данных как залог результативного внедрения аналитических процессов в деятельность компании.

2.1.2 Аналитическая культура и организационные структуры

В исследовании компании McKinsey о внедрении аналитических практик в современных компаниях было выявлено, что для построения эффективно функционирующей аналитической экосистемы в организации должны выполняться не только технические предпосылки, но и культурные. В связи с этим принято говорить об аналитической культуре компании [26].

Аналитическая культура (от англ. Analytical Culture), как правило, формируется по мере развития экосистемы управления данными компании и

внедрения результатов аналитической работы в операционную и стратегическую деятельность. Аналитическая культура – это совокупность наиболее эффективных и устойчивых форм управления данными, выработанных в процессе адаптации организации к внутренним и внешним условиям [27].

Современные руководители рассматривают аналитическую культуру своей организации в качестве стратегического инструмента, позволяющего ориентировать все подразделения и определенных лиц на общие цели, мобилизовать инициативу сотрудников, связанную с внедрением данных в бизнес-процессы и механизмы принятия решений.

Развитая аналитическая культура компании позволяет опираться в большей степени не на накопленный опыт и авторитет определенных лиц, принимающих решения, а на выводы, в основе которых лежит фактическая информация. Для развития аналитической культуры необходима глубокая ценностная трансформация компании, предусматривающая вовлечение как менеджмента, так и лидеров внедрения аналитических практик в компании.

Основными аспектами аналитической культуры являются следующие положения [28]:

- ориентация на данные при принятии решений;
- учет всей полноты доступной информации при принятии решений;
- формирование дата-центрированного (от англ. data-centric) мышления сотрудников компании;
- открытость данных и предоставление максимально широкого спектра данных всем сотрудникам компании, а также обеспечение обмена данными;
- умеренный скептицизм по отношению к идеям и тестирование бизнес-гипотез на основе данных;
- ориентация на бизнес-результат и извлечение ценности для компании;

Интенсивное внедрение технологий обработки больших массивов данных и повышение стратегической значимости данных приводит к необходимости поиска способов построения эффективной аналитической культуры.

Формирование аналитической культуры представляет собой эволюционный процесс, в рамках которого происходит переход от более простых аналитических практик к более сложным. Накопленный опыт в реализации аналитических проектов, развитость инфраструктуры сбора и обработки данных, а также количество ресурсов, направляемых компанией на совершенствование аналитической экосистемы являются ключевыми факторами, определяющими динамику развития аналитической культуры.

Также развитие аналитической экосистемы в телекоммуникационных компаниях происходит эволюционно. Так сначала происходит формирование систем отчетности, далее формируется экосистема данных, после чего происходит формирование систем управления мастер-данными, и, когда система устойчиво работает, она отлажена и данные пользуются доверием у

бизнес-пользователей, происходит переход к углубленной аналитике на основе систем машинного обучения.

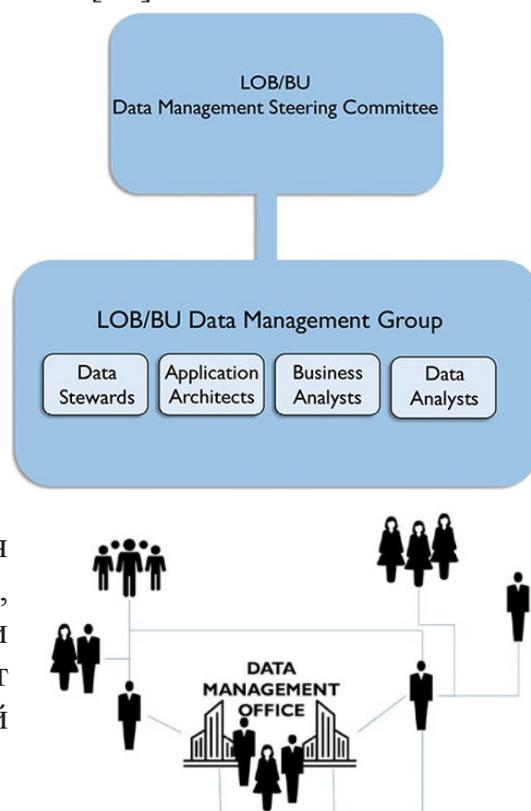
Драйвером развития аналитической культуры в телекоммуникационных компаниях, как правило, является менеджмент компании. Однако, на фоне активного внедрения аналитических практик в бизнес-процессы компаниями индустрии, появляется необходимость в развитии аналитической культуры на каждом уровне управленческой иерархии. В связи с этим фактором, а также с появлением регуляторных требований, регламентирующих процессы управления данными, в телекоммуникационных компаниях начинают возникать новые должности и бизнес-роли.

Например, во многих телекоммуникационных компаниях, назначаются главные специалисты по данным (от англ. Chief Data Officer), главной задачей которых является формирование стратегии управления данными компании. В обязанности CDO также входят: бюджетирование ресурсов для развития аналитической экосистемы компании, внедрение передовых аналитических практик, контроль соблюдения принятых практик управления данными, а также консультирование бизнеса в рамках проектов, связанных с интенсивным использованием данных.

По мере развития и усложнения аналитической культуры в компании, становится необходимым структурирование процессов управления данными, а также построение организационной схемы компании, отвечающей условиям оптимальности и эффективности управления жизненным циклом данных.

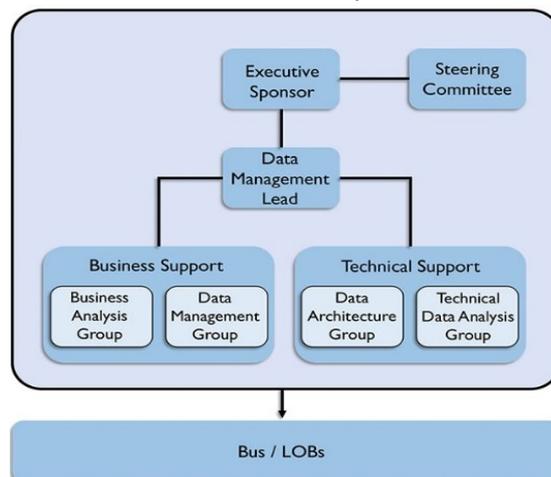
В соответствии с методологией DMBOOK выделяются 5 типов организационных моделей управления данными [24]:

- **Децентрализованная операционная модель** – обязанности по управлению данными распределены между различными подразделениями компании. В данной управленческой модели практикуется подход снизу-вверх, когда отдельные инициативы по управлению данными унифицируются и складываются в единую систему на более высоких уровнях управленческой иерархии;
- **Сетевая операционная модель** – в этой модели обязанности управления данными зависят от связей, выстраиваемых между инициативными группами. Такая модель позволяет надстраивать новый слой

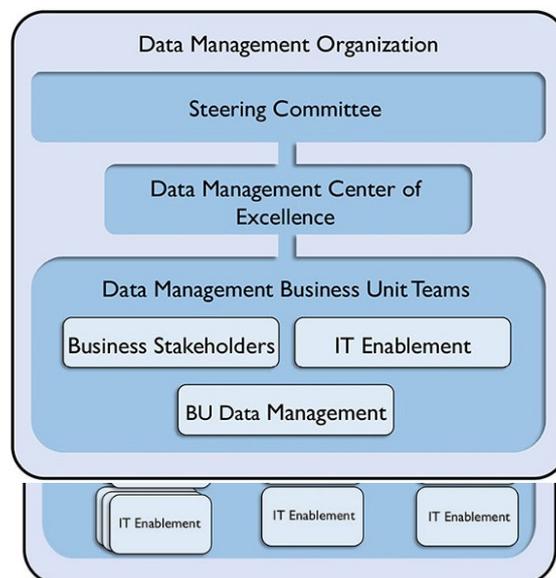


организационной модели, практически не влияя на остальные;

- **Централизованная операционная модель** – одна из наиболее устоявшихся и формальных организационных моделей. Данная модель представляет собой ступенчатую субординационную схему, в рамках которой решения по управлению данными принимаются «сверху-вниз». Как правило, ключевой фигурой, в централизованной модели, является CDO;



- **Гибридная операционная модель** – в соответствии с названием, данная модель предусматривает использование элементов подходов как централизованной, так и децентрализованной организационных моделей. В гибридной модели централизованная структура управления данными взаимодействует с децентрализованной. Например, представители отдельных инициативных групп или подразделений могут входить в комитет, результаты работы которого рассматриваются на верхних управленческих уровнях;



- **Федеративная операционная модель** – вариация гибридной модели, которая предусматривает дополнительные централизованные или децентрализованные слои в организационной схеме, что часто востребовано в крупных международных корпорациях с высоким уровнем унификации бизнес-процессов среди подразделений;

В связи с одновременной сложностью механизмов управления данными в телекоммуникационных компаниях и необходимостью в согласованной политике, наиболее оптимальной является гибридная модель. Этот тип организационных моделей обладает множеством преимуществ. Главное из них выражается в том, что процесс принятия решений остается децентрализованным, но при этом соотносится с общей политикой управления данными компании посредством руководящего комитета.

Таким образом, рассмотренная выше гибридная организационная модель по методологии DMВОК, позволяет органично внедрить аналитическую культуру в корпоративную культуру компании.

2.1.3 Итерационные методологии анализа данных

В связи с тем, что в рамках аналитических проектов часто требуется некоторое количество стандартизированных операций, появляется множество подходов, направленных на структурирование и унификацию аналитических процессов.

Согласно опросу, проведенному ресурсом KDnuggets в 2014 году, наиболее популярны три из них [29]:

- SEMMA;
- KDD;
- CRISP-DM;

SEMMA

SEMMA – аббревиатура от английских слов Sample, Explore, Modify, Model, Assess (соответственно Выборка, Изучение, Преобразование, Моделирование, Оценка) – общая методология и последовательность шагов интеллектуального анализа данных, которая была предложена американской компанией SAS, одним из крупнейших производителей статистического ПО и анализа данных. В отличие от CRISP-DM методология SEMMA фокусируется на аспектах моделирования без учета бизнес-компоненты [30]. Однако, точно также, как и CRISP-DM данный подход позиционируется как межотраслевой подход к последовательному интеллектуальному анализу данных.

Методология SEMMA состоит из пяти этапов, которые представлены на циклической диаграмме (Рисунок 2.3).

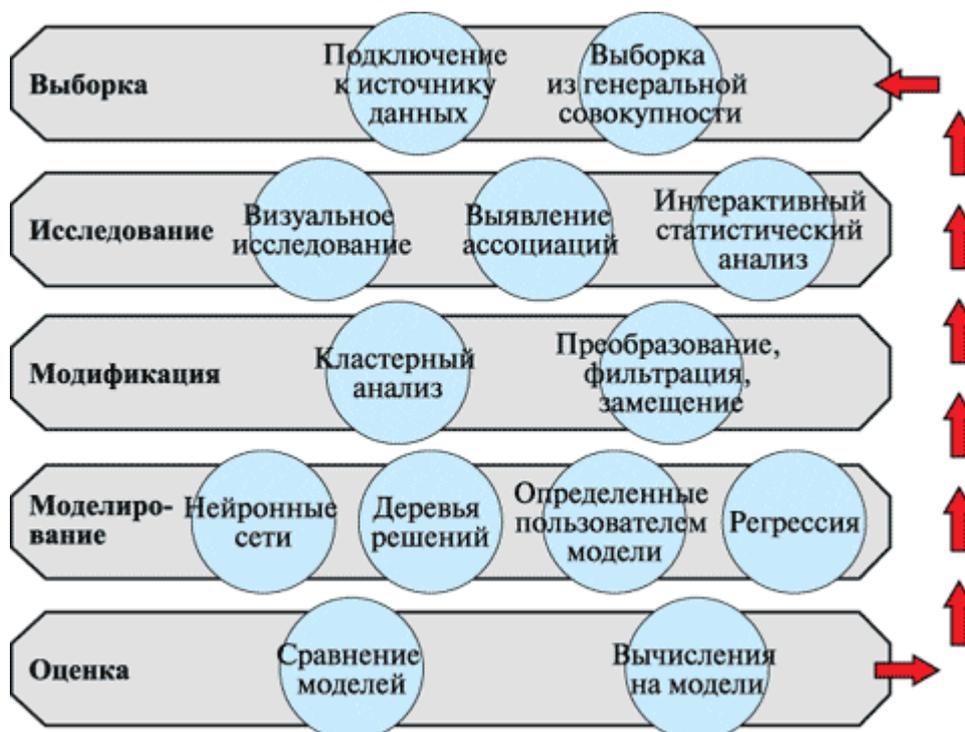


Рисунок 2.3 – Диаграмма компонентов методологии SEMMA

Примечание – Источник: SEMMA (<https://www.bigdataschool.ru/wiki/semma>)

Данная методология интегрирована в пакет анализа данных SAS Data Mining Solution [31]. SEMMA затрагивает только технические аспекты анализа данных, что является существенным недостатком при выборе методологии анализа данных. Таким образом, CRISP-DM является более полной методологией, учитывающей бизнес-контекст решаемой аналитической задачи.

KDD Process

KDD (от англ. knowledge discovery in databases) или обнаружение знаний в базах данных. Также под KDD понимают процесс поиска полезных знаний в «сырых» данных. KDD состоит из следующих этапов:

- Подготовка данных;
- Предобработка данных;
- Трансформация и нормализация данных;
- Data Mining;
- Постобработка данных;

KDD определяет последовательность действий, которую необходимо выполнить для получения знаний из исходных данных. Под «знаниями» подразумевается некоторое абстрактное понимание ценности данных. KDD является одной из наиболее применимых методологий углубленной аналитики данных. Схема процесса обработки данных по методологии KDD ниже (Рисунок 2.4).

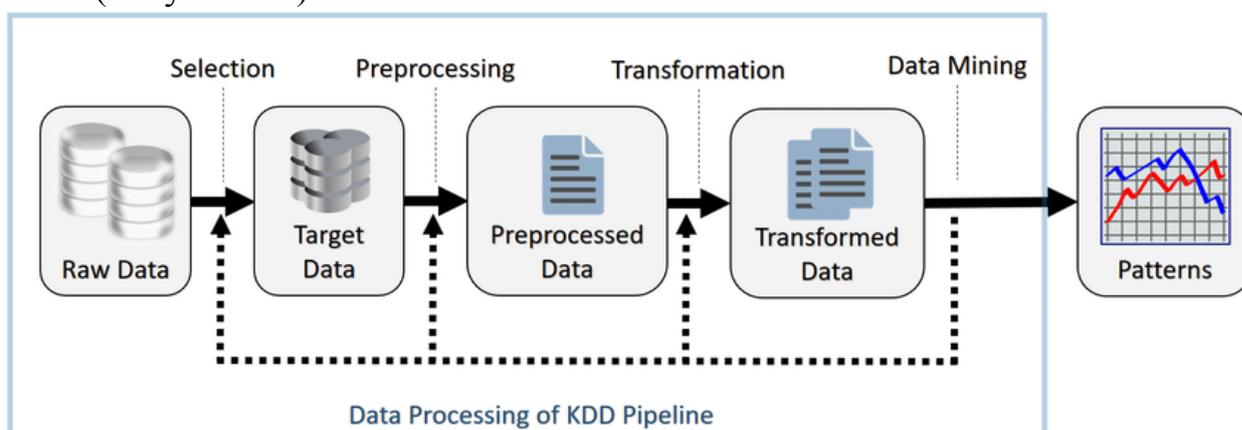


Рисунок 2.4 – Диаграмма процесса KDD

Примечание – Источник: KDD (<https://basegroup.ru/community/articles/kdd>)

KDD является менее проработанной методологией и не имеет важного компонента моделирования данных, часто требующегося для углубленной аналитики [32].

CRISP-DM

Рассмотрим методологию CRISP-DM, получившую наибольшее распространение при решении аналитических задач в различных отраслях. CRISP-DM – это межотраслевой стандарт углубленной аналитики данных (от англ. CRoss Industry Standard Process for Data Mining), разработанный консорциумом ESPRIT в 1999 году. CRISP-DM описывает общие процессы и подходы к аналитике данных, используемые в промышленных data-mining проектах независимо от конкретной задачи или индустрии. В связи с большой популярностью данной методологии среди проектов углубленной аналитики данных, рассмотрим ее более детально.

Согласно обзору методологий data-mining, CRISP-DM получил наибольшее распространение в связи множеством преимуществ по отношению к другим методологиям. Среди них необходимо выделить отсутствие ограничений по применению методологии в различных индустриях, а также отсутствие технических и платформенных ограничений [33]. CRISP-DM интегрирована в аналитический продукт IBM SPSS Modeler, что позволяет реализовать end-to-end аналитическое решение в единой экосистеме.

Модель CRISP-DM является моделью жизненного цикла исследования данных. Цикл состоит из 6 фаз. Ниже представлена диаграмма CRISP-DM (Рисунок 2.5) [34].

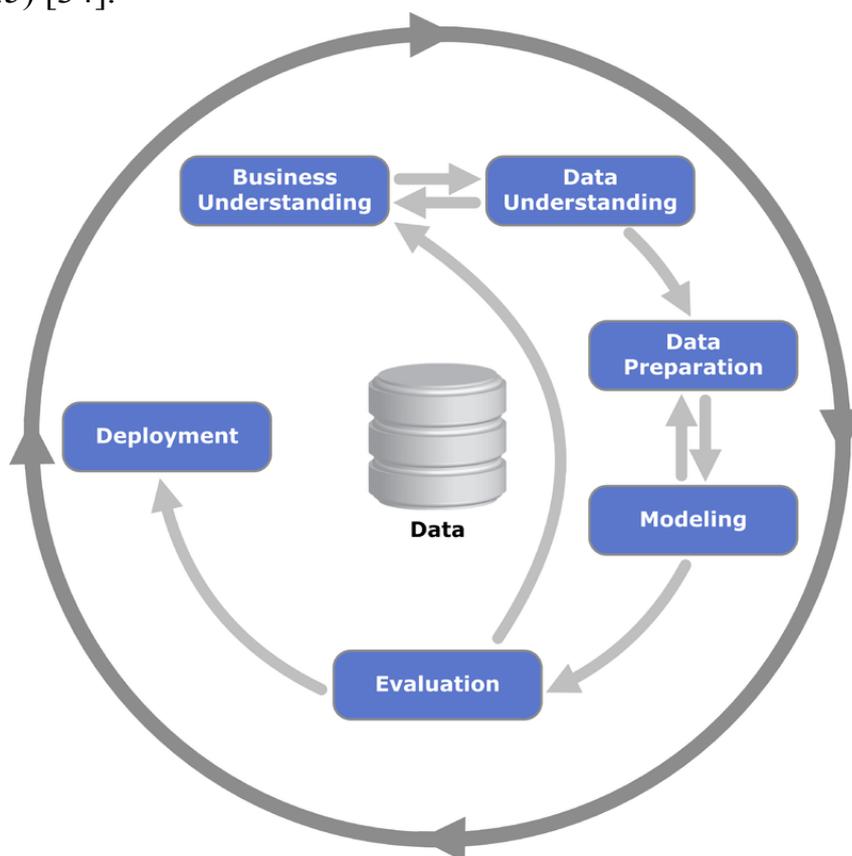


Рисунок 2.5 – Диаграмма аналитического цикла CRISP-DM

Примечание

–

Источник:

CRISP-DM

(https://www.ibm.com/support/knowledgecenter/en/SS3RA7_17.1.0/modeler_crispdm_ddita/modeler_crispdm_ddita-gentopic1.html)

Методология CRISP-DM также является достаточно гибкой и позволяет возвращаться на предыдущие фазы, что является ценным фактором при реализации аналитических проектов. CRISP-DM отлично подходит для проектов углубленной аналитики, в рамках которых используются модели машинного обучения и техники статистического моделирования [34].

Рассмотрим 6 фаз модели исследования данных:

1. **Бизнес-анализ.** На первой фазе происходит бизнес-анализ, постановка целей проекта, оценивается текущая ситуация и наличие необходимых для реализации проекта ресурсов. Помимо этого, на первой фазе оцениваются потенциальные риски и способы их минимизации. Также часто практикуется построение ROI-оценки для определения ценности проекта. Как правило, ROI получается оценить только после этапа пилотного моделирования. После постановки задачи в бизнес-терминах и оценки привлекательности реализации проекта, происходит формализация аналитической задачи и ее описание в технических терминах. В частности, выбираются метрики оценки качества и определяются критерии успешности реализации проекта. После того, как задача формализована, ясны цели проекта и его ценность, составляется план, содержащий детальное описание всех шести фаз внедрения.
2. **Анализ данных.** В рамках данной фазы происходит первичная оценка данных, определение их качества и достаточности. На данной фазе применяется описательная аналитика с целью выявления закономерностей в данных, поиска потенциала данных и определение оптимальных стратегий предобработки данных для следующих фаз цикла CRISP-DM. Первым шагом второй фазы является сбор данных и анализ источников данных. Далее составляется описание данных и модели данных. После этого происходит оценка качества данных на предмет наличия пропусков, незаполненных значений, ошибок в данных, а также выявляется неконсистентность кодировки данных. Финальным шагом фазы анализа данных является формирование отчета с табличными материалами и визуализацией. Результат данной фазы – понимание потенциала данных, с учетом их качества и описательных свойств.
3. **Подготовка данных.** Как правило, данная фаза CRISP-DM является наиболее продолжительной для большинства проектов и занимает от 50% до 70% времени реализации проекта, что обусловлено ее трудоемкостью. Первый шаг третьей фазы – отбор данных, наиболее релевантных для решения аналитической задачи, а также удовлетворяющих требованиям используемых техник моделирования. Затем происходит очистка данных и от пропущенных значений и ошибок, выявленных на второй фазе. Также часто применяются

техники генерации данных (от англ. feature engineering), позволяющие сформировать качественные признаки, отвечающие целям моделирования лучше, чем базовый набор данных. При реализации аналитических проектов, часто необходимо интегрировать данные из различных внутренних и внешних источников. Под интеграцией подразумеваются процедуры агрегации, объединения и соединения данных. Результатом третьей фазы является представление данных в табличном виде, пригодном для использования в аналитическом ПО.

4. **Моделирование.** На четвертой фазе начинается процесс моделирования данных, который, как правило, является итерационным: тестируется множество моделей, отбираются наиболее эффективные по результатам оценки их качества, а также происходит настройка параметров моделей. Важным этапом фазы моделирования является выбор алгоритмов, который зависит от природы данных и специфики решаемой задачи. Далее происходит построение плана тестирования моделей, определяются стратегии моделирования данных и определения метрик их качества. После этого происходит обучения модели и фиксация результатов. В результате моделирования, формируется пул моделей с оцененными метриками качества. Производится техническая оценка метрик качества, соответствуют ли результаты моделирования поставленным требованиям из первой фазы CRISP-DM, а также оценивается полнота достижения бизнес-целей. Если критерии успеха не выполняются, третья и четвертая фазы повторяются заново. Результатом четвертой фазы является построенная модель, отвечающая требованиям качества, а также приобретенное понимание данных и закономерностей в них.
5. **Оценка результата.** На пятой фазе CRISP-DM оценивается соответствие полученных результатов на этапе моделирования бизнес-целям проекта. Также на данной фазе происходит интерпретация результатов в терминах экономической выгоды и бизнес-результатов. Важным шагом пятой фазы является анализ хода выполнения проекта, а также выявление сильных и слабых сторон примененных подходов для оптимизации реализации последующих аналитических проектов. Результатом пятой фазы является принятие решений касательно внедрения результатов аналитического проекта.
6. **Внедрение.** На последней фазе аналитического цикла происходит поставка моделей и результатов по оговоренным на этапе планирования способам. Шестая фаза предусматривает развертывание модели и ее интеграцию в информационную систему компании. Также на финальном этапе цикла может быть инициировано изменение бизнес-процессов компании. Определяются механизмы мониторинга и поддержки внедряемых моделей и аналитических продуктов. Результатом шестой фазы является формирование отчета по итогам

первых пяти фаз цикла, который презентуется заинтересованным лицам и бизнес-пользователям.

Этапы цикла CRISP-DM представлены в таблице 2.1.

Таблица 2.1 – Этапы цикла CRISP-DM

Бизнес-анализ	Анализ данных	Подготовка данных	Моделирование	Оценка результата	Внедрение
Определение бизнес-целей	Сбор данных	Выборка данных	Выбор алгоритмов	Оценка результатов	Внедрение
Оценка текущей ситуации	Описание данных	Очистка данных	Подготовка плана тестирования	Оценка процесса	Планирование мониторинга и поддержки
Определение целей аналитики	Изучение данных	Генерация данных	Обучение моделей	Определение следующих шагов	Подготовка отчета
Подготовка плана проекта	Проверка качества данных	Интеграция и форматирование данных	Оценка качества моделей	Интерпретация результатов	Обзор проекта

Примечание – Разработка автора

В телекоммуникационной сфере данная методология подходит для применения в различных типах аналитических проектов. Для Ad hoc аналитики CRISP-DM позволяет в соответствии с четко очерченной структурой быстро найти оптимальное решение. Для моделей, выводимых в постоянную работу (от англ. production models) CRISP-DM позволяет структурированно подойти ко всем этапам сбора, обработки, анализа и моделирования данных и нивелировать ошибки, которые могут возникнуть на каждом из этапов. Также CRISP-DM удобно интегрируется в бизнес-модели DAaaS, где требуется высокая скорость построения аналитических решений, а также извлечение максимальной ценности из имеющихся данных. Унификация аналитических процессов позволяет достичь более высокой отдачи от инвестиций (ROI) проекта.

В отличие от SEMMA и KDD методология CRISP-DM позволяет структурированно подойти к построению аналитических процессов, учитывающих как ранние фазы проекта, связанные со сбором требований и первичной оценкой проекта до финального этапа внедрения. Таким образом, CRISP-DM наиболее полно соответствует требованиям рынка телекоммуникаций и может быть рекомендована для внедрения в качестве элемента аналитической экосистемы.

Тем не менее, в CRISP-DM отсутствуют компоненты управления проектами и возможность интеграции agile-методологий, а также не рассматривается важный для многих проектов этап прототипирования аналитических решений, поэтому в 2015 году компанией IBM был разработан подход ASUM-DM (от англ. Analytics Solutions Unified Method for Data Mining and predictive analytics), учитывающий потребности современных проектов, в которых применяются технологии анализа больших данных и машинного обучения. ASUM-DM является своеобразной надстройкой CRISP-DM. ASUM-DM учитывает различные аспекты командной работы над

аналитическими проектами, повышая эффективность взаимодействия аналитиков [35].

Таким образом, в соответствии с проведенным выше анализом методологий построения систем управления данными, организационных схем, а также методологий ведения аналитических проектов, можно сделать выводы касательно их ценности с точки зрения внедрения в аналитическую экосистему телекоммуникационных компаний. Также важным является возможность интеграции выявленных лучших практик между собой.

С точки зрения эффективности управления данными в компании, наибольшую ценность представляет методология DAMA DMBOK, позволяющая описывать процессы, связанные с управлением данными на различных уровнях абстракции. Помимо этого, ценность методологии заключается в удобстве анализа различных компонентов системы управления данными в контексте бизнес-процессов и бизнес-результатов.

При рассмотрении организационных моделей, наиболее перспективной представляется гибридная модель, сочетающая в себе лучшие практики как централизованной, так и децентрализованной моделей.

Для ведения аналитических проектов, наибольшую ценность имеют две взаимосвязанные методологии – CRISP-DM и ASUM-DM. Они позволяют декомпозировать процесс построения аналитических решений на логические блоки, что оказывает положительный эффект на качество, скорость и инвестиционную привлекательность реализуемых аналитических проектов.

2.2 Классификация типов данных телекоммуникационных компаний

Современные телекоммуникационные компании стремятся к принятию наиболее оптимальных решений на основании всей имеющейся информации в режиме реального времени, при этом минимизируя издержки на обеспечение аналитических систем. Поэтому для реализации потенциала имеющихся данных, телекоммуникационные компании нуждаются в стабильно работающей аналитической экосистеме данных, которая выполняет функцию администрирования процессов обработки данных.

Компания McKinsey построила оценку доли прироста прибыли телекоммуникационных компаний, обеспечиваемую успешным внедрением технологии обработки больших данных и внедрением новых аналитических решений. На графике ниже представлена оценка влияния внедряемых решений на прибыль среди телекоммуникационных компаний, инвестировавших в Big Data в 2015 году (Рисунок 2.6) [36].

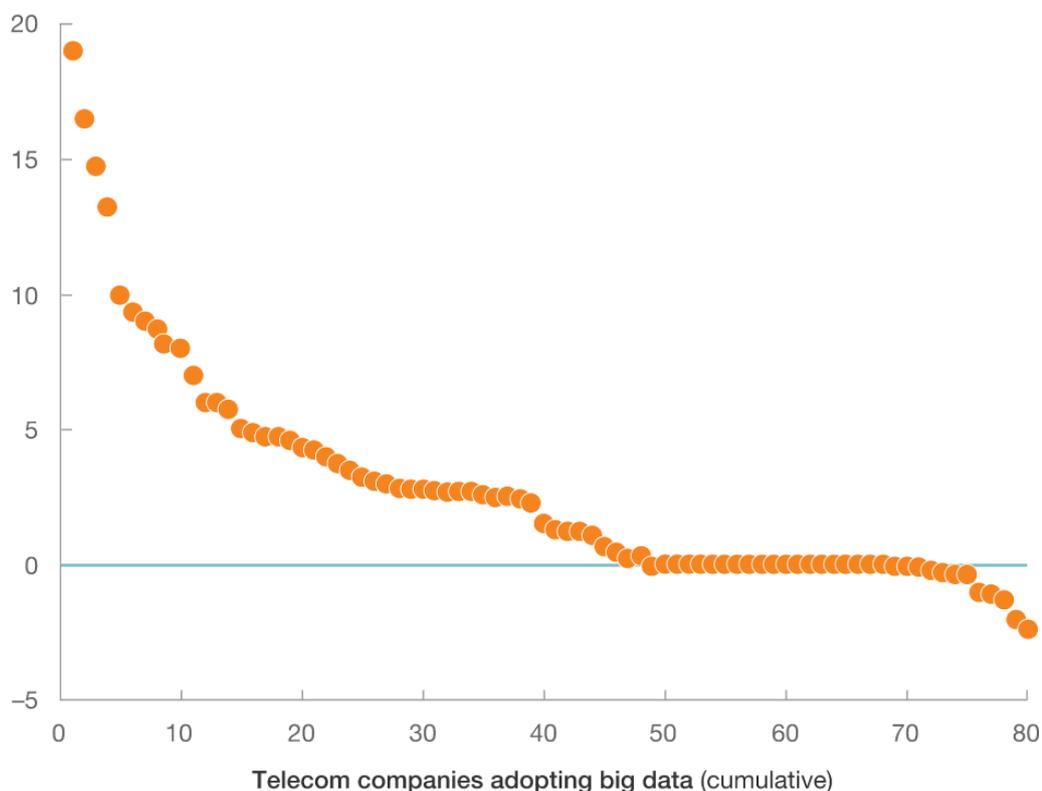


Рисунок 2.6 – Эффект от внедрения телекоммуникационными компаниями решений по аналитике больших данных

Примечание – Источник: McKinsey

<https://www.mckinsey.com/industries/telecommunications/our-insights/telcos-the-untapped-promise-of-big-data>

Для понимания ценности и применимости данных в различных бизнес-контекстах необходима их классификация. Данные разделяют на **внутренние** и **внешние**. Внешние данные - данные, получаемые компаниями из различных внешних источников.

Примерами внешних данных являются открытые источники, такие как гео-данные карт (применимы для геокодирования адресного пространства, определения местоположения организаций и др.), информация из социальных сетей (применимы для анализа профиля интересов, активности пользователей социальных сетей, построения социального графа и др.), информация, извлекаемая из медиапространства (применимы для анализа новостей, проведения маркетинговых исследований, анализа восприятия бренда и др.). Внешние данные имеют сравнительно более низкую степень

надежности, так как для их использования и построения валидных выводов необходима дополнительная верификация.

Внутренние источники данных представляют собой все накапливаемые организацией данные, которые генерируются целенаправленно или в качестве побочного информационного продукта в результате функционирования бизнеса. Они представляют собой наибольшую экономическую ценность и выступают в роли одного из наиболее ценных нематериальных активов телекоммуникационных компаний. Внутренние данные создаются в рамках различных корпоративных информационных систем, таких как биллинговые, CRM, BPM, ERP, HRM и других систем. Помимо этого, наиболее релевантные данные, используемые для аналитики, хранятся ВІ-системами в преобразованном виде и должны соответствовать требованиям доступности, достоверности и полноты.

Также данные принято классифицировать согласно регулярности их генерации - это могут быть **регулярные** и **нерегулярные** данные. Регулярные данные, такие как информация о потребляемых клиентами услугах и оплачиваемых счетах, как правило, накапливаются за весь доступный исторический период. Нерегулярные данные хранятся за ограниченный исторический период в транзакционном виде.

Данные также делятся на три типа в зависимости от степени их структурированности: **неструктурированные** (данные, у которых нет формально определенной и описанной структуры), **полуструктурированные** (данные, не имеющие определенной схемы, но имеющие определенную единообразную структуру), **структурированные** (данные, имеющие формально определенную схему и структуру). В сфере телекоммуникаций, в аналитических целях, как правило, используются структурированные внутренние данные, реже - полуструктурированные и неструктурированные.

Для описания схем взаимодействия между различными типами данных и различными классами информационных систем компаниями используются **метаданные**, **мастер-данные** и **иерархические данные**. Метаданные - данные, описывающие конфигурацию других данных. Мастер-данные - данные, содержащие ключевую информацию о функционировании бизнеса. Иерархические данные - данные, в которых хранятся отношения и взаимодействия между данными.

Также в контексте решения аналитических задач, данные подразделяются на **обработанные** и **необработанные**. Необработанные или «сырые» данные представляют собой результат первичного накопления данных о некоторых событиях, регистрируемых информационными системами телекоммуникационной компании. В необработанных данных, в отличие от обработанных, не происходит наложения бизнес-логики. После обработки данных зарегистрированное событие интерпретируется посредством правил, определенных бизнес-логикой компании.

Для формирования представления о том, какие возможности открываются перед телекоммуникационными компаниями, использующими

технологии обработки больших данных, необходимо изучить тип и структуру данных, генерируемую телеком-компаниями.

Существует множество классификаций данных, имеющихся у телекоммуникационных компаний. Компания McKinsey в своем отчете о перспективах развития технологий обработки больших данных в индустрии телекоммуникаций выделяет несколько типов пользовательских данных, собираемых компаниями. Представляется целесообразным обобщить их в 5 групп [36]:

1. Данные о клиентах;
2. Данные об использовании услуг;
3. Гео-данные и гео-информационные системы в индустрии телекоммуникаций;
4. Информация из каналов заботы о клиенте;
5. Данные о взаимодействии клиентов;

Для оценки потенциала применения различных типов данных, имеющихся у телекоммуникационных компаний, необходим анализ их потенциала и точек приложения.

2.2.1 Данные о клиентах

Телекоммуникационные компании накапливают значительные массивы персональной клиентской информации. Персональная информация является основным источником данных для решения аналитических задач, так как именно она наиболее полно соответствует требованиям качества, в связи с ее использованием для однозначной идентификации клиентов при предоставлении услуг, а также требованиями регуляторов. К этому типу данных относятся социо-демографическая информация о клиенте, например, пол, возраст, регион и персональные идентификаторы клиента. Также в случае сбора телеком-компаниями информации о месте проживания или адреса почтовой доставки, информация о клиентах обогащается адресным пространством, которое можно отнести к гео-данным.

Источники данных о клиентах могут быть как внутренние, так и внешние. Например, некоторые телеком-компании анализируют поведенческий профиль на основании данных из открытых источников (социальные сети), а также государственных данных (кредитный рейтинг).

2.2.2 Данные об использовании услуг

Данные об использовании услуг – телекоммуникационные компании собирают информацию о типе и количестве используемых клиентом сервисов, интенсивности их использования, данные о потреблении трафика и частоте использования услуг. Большая часть информации, генерируемой в телеком-компаниях, является информацией, создаваемой корпоративными

информационными системами (КИС) в автоматическом режиме. Они являются неотъемлемой частью программной инфраструктуры любой телеком-компания и выполняют критически важные функции для управления компанией.

Основным источником данных об использовании услуг являются CDR - детализированная информация о звонках из CDR-систем (Call Detail Record). Детализированные записи о вызовах являются ценным источником информации, который открывает новые возможности для телекоммуникационных компаний по максимизации прибыли. Разнообразие данных и высокая интенсивность генерации данных позволяет отнести CDR-источники к классу больших данных. В связи с этим, как правило, отдельно выделяют CDR-аналитику [37].

CDR-записи содержат детализированную информацию в рамках звонковой или интернет-сессии. Например, в CDR сохраняется полная информация о транзакции в сети телеком-оператора: информация о начале и конце звонка, длительности, участниках звонка (номера и системные идентификаторы), идентификаторы сот, а также информация о посещаемых сайтах в рамках интернет-сессий. Информация из CDR имеет высокий уровень детализации и редко используется в аналитических напрямую в связи с большим размером CDR-источников. CDR используется телеком-компаниями для реализации базовых бизнес-процессов, необходимых для функционирования бизнеса, таких как: расчетные системы, биллинговые системы и системы управления эффективностью сети и других, однако CDR-информация также участвуют при решении практически всех аналитических задач.

Примеры использования CDR-данных в аналитических целях как телеком-компаниями, так и компаниями из других индустрий, представлены в таблице 2.2.

Таблица 2.2 – Примеры использования данных об использовании услуг в телекоммуникационных компаниях и других сферах

В рамках телекоммуникационных компаний	В других сферах
Аналитика в режиме реального времени, а также использование системах оперативного принятия решений (DMS);	Интеллектуальный маркетинг;
Точный маркетинг (Precise Marketing): оптимизация предложений клиентам, идентификация оттока, персонализация предложений;	Политические кампании;
Повышение операционной эффективности;	Социальные исследования;
Улучшение клиентского опыта;	Мероприятия по планированию

	городской среды;
Анализ прибыльности сервисов;	Построение социального графа;
Поведенческий анализ базы;	Системы общественной безопасности и системы реагирования на чрезвычайные происшествия;

Примечание – Разработка автора

2.2.3 Гео-данные и гео-информационные системы в индустрии телекоммуникаций

На данный момент во многих телекоммуникационных компаниях наблюдается бурный рост использования гео-информации в системах поддержки принятия решений [38]. Она используется как для анализа абонентской базы, так и для анализа процесса эксплуатации телекоммуникационной инфраструктуры. Фактически, гео-данные, используемые в телекоммуникационных компаниях, извлекаются из CDR-источников и являются производной информацией из CDR. CDR-записи содержат информацию об идентификаторах сот и при объединении этих данных с данными об их физическом местоположении, возникает возможность геолоцирования абонентов на основании их активности в сети.

Результатом построения гео-аналитической картины динамики абонентской базы становится возможность реализации гео-маркетинговых кампаний и рекламного таргетинга абонентов по географическому признаку. Гео-информация позволяет осуществлять более эффективный маркетинговый анализ и проводить углубленный конкурентный анализ.

Существует множество контекстов применения гео-информации телеком-компаниями [39]. Она может применяться для:

- рыночной сегментации абонентской базы;
- оценки потенциала развертывания сети в определенных географических зонах;
- выделения зон повышенной концентрации абонентов;
- оптимизации сетевой инфраструктуры и управления емкостью сети;
- проецирования статистической картины для зон без покрытия инфраструктурой оператора;
- мониторинг эффективности каналов дистрибуции;
- повышения эффективности клиентского обслуживания и планирования мероприятий по увеличению рыночного проникновения.

На основании поведенческого анализа могут быть определены базовые станции, на которые приходится максимальная событийная активность абонента (наибольшее количество входящих и исходящих звонковых

событий, СМС, интернет-сессий). Также при применении эвристических правил могут быть определены места проживания или места работы абонента. Такой подход дает информацию в большом приближении - он не позволяет сделать точных выводов о местоположении абонента, так как определяется местоположение базовой станции, а не местоположение терминального оборудования. Тем не менее, при дополнении этой информации картографическими данными и информацией из корпоративных и открытых источников, можно получить более ценную и репрезентативную картину. Гео-информации, основанной на информации о местоположениях базовых станций, как правило, достаточно для выделения паттернов геомобильности, однако велика доля погрешности, определяемая радиусом зоны покрытия соты на базовой станции.

Преобразованные гео-данные также используются телекоммуникационными компаниями в качестве предикторов для углубленной аналитики с использованием методов статистического моделирования. Как правило, гео-предикторы вносят существенный вклад в качество моделей. Конструирование гео-предикторов для различных задач углубленной аналитики зависит от специфики решаемой задачи.

Развитие B2B консалтинга в телеком-компаниях также позволяет находить применение гео-информации в других сферах, в виде предлагаемых заказчикам data-продуктов. Такие решения, основанные на гео-информации телеком-компаний, например, могут быть связаны с оптимизацией работы розничных сетей крупного бизнеса, размещением объектов торговой инфраструктуры, а также размещением наружной рекламы, ориентированной на определенный клиентский сегмент. В сферах, в которых требуется повышенная точность идентификации, гео-информация используется для точной идентификации личности по признакам гео-мобильности.

В сфере гео-аналитики в течение ближайших лет ожидаются коренные изменения, связанные с масштабным внедрением геолокационных решений на базе технологии триангуляции, позволяющей более точно определять местоположение радиопередатчика (с погрешностью до 50 метров). Триангуляция - это процесс, с помощью которого местоположение радиопередатчика может быть определено путем измерения радиального расстояния или направления принятого сигнала от двух или трех разных точек. Таким образом, станет возможным построение footfall-аналитики или точной траектории движения пользователей, что значительно увеличит аналитическую ценность гео-информационных решений.

Гео-информация как в агрегированном виде, так и детализированная на уровне отдельного абонента, представляет собой большую ценность для телеком-компаний, так как дает пространственное измерение для анализа и позволяет применять эту информацию для решения большого класса аналитических задач. Развитие геоинформационных сервисов и внедрение гео-аналитических решений телекоммуникационными компаниями позволяет

снижать издержки на обслуживание инфраструктуры, оценивать точки роста, а также оптимизировать работу дистрибьюторской сети.

2.2.4 Информация из каналов заботы о клиенте

Такой тип информации создается в результате взаимодействия клиентов с телеком-оператором. Согласно исследованию, проведенному Ernst & Young, 68% телекоммуникационных операторов оценивают управление клиентским опытом как ключевой приоритет развития компании [9].

По оценкам экспертов, данный тип информации практически не монетизируется телекоммуникационными компаниями, в то время как именно он содержит ценную обратную связь касательно причин оттока абонентов. Также такая информация позволяет определить неэффективно функционирующие объекты дистрибьюторской сети и предпринять необходимые действия по их оптимизации.

К информации из каналов заботы о клиенте относится контактная информация обратившегося, содержание обращения, результат обращения, вносимый оператором, также звонок может храниться в виде аудиозаписи или транскрибироваться в текст. В последнее время с ростом популярности использования цифровых каналов взаимодействия, таких как чаты с операторами посредством мессенджеров, появляется новый источник хорошо структурированной информации, которая может быть использована в аналитических целях.

Развитие подходов моделирования, связанных с генерацией текста позволяет внедрять системы самообслуживания на базе чат-ботов, способных поддерживать осмысленный диалог и находить необходимую для обратившегося информацию.

С появлением высоко эффективных моделей глубокого обучения и совершенствованием подходов обработки естественного языка на базе искусственных нейронных сетей, появилась возможность обработки аудиозаписей обращений клиентов в автоматическом режиме. Для этого, как правило, требуется перевод аудио-звонка в текст посредством моделей speech2text, далее текст обрабатывается языковыми моделями. Массивы записанных звонков могут быть проанализированы алгоритмом для определения метрик качества работы оператора, соблюдения оператором регламентированных стандартов работы с клиентом, определения эмоционального состояния клиента, оказания помощи в реальном времени сотрудникам call-центра. Также на базе распознанной в текстовый формат речи могут быть сняты различные метрики звонка, такие как количество перебиваний, время ожидания ответа, процент молчания, использование нецензурной лексики, попытки кросс-продаж, упоминания названий продуктов или названий компаний-конкурентов и множество других.

На базе накопленных исторических данных из каналов заботы о клиенте, могут быть предприняты действия по предотвращению конфликтных ситуаций и ситуаций, причиняющих неудобства клиенту. Таким образом, внедрение аналитических решений, использующих данные из каналов заботы о клиенте, позволяет снизить издержки на сопровождение клиентов, улучшить клиентский опыт, сократить количество ручной работы, направленной на контроль качества и соблюдение регламента взаимодействия с клиентами, а также повысить общее коммуникации с клиентом [40].

2.2.5 Данные о взаимодействии клиентов

На основании детализированной информации звонков (CDR-данных) и персональных данных абонентов телеком-компаниями анализируются взаимодействия между абонентами. Задачи анализа взаимодействий абонентов формализуются при помощи теории анализа социальных сетей (от англ. social network analysis theory - SNA), которая представляет собой одну из ключевых инструментальных техник современной социологии и широко применяется телекоммуникационными компаниями в маркетинге, бизнес-анализе, а также при решении задач поведенческого моделирования [41]. Отдельные участники сети отображаются в виде узлов, а отношения между участниками отображаются в виде связей.

Такое представление позволяет оценивать множество метрик, описывающих социальную динамику абонентской базы:

- **степень гомогенности группы:** степень, с которой участники сети формируют связи между собой по сравнению с несхожими;
- **множественность группы:** количество форм, содержащихся в связи;
- **взаимность группы:** степень, с которой двое участников отвечают друг другу взаимностью в рамках взаимодействий;
- **степень закрытости сети:** мера полноты реляционных триад, то есть факт того, что узлы, связанные с отдельным узлом, также будут связаны и между собой;
- **соседство:** склонность участников иметь больше связей с участниками, расположенными ближе географически;
- **мера плотности сети:** отношение количества связей сети к общему возможному количеству связей;
- **выявление структурных пробелов:** выявление пробелов между несколькими частями социальной сети;

Телекоммуникационными компаниями SNA применяется в коммуникативных исследованиях абонентской базы. SNA позволяет телеком-компаниям понять, каким образом динамика абонентской базы определяется поведением отдельных индивидов и малых групп, а также, как поведение отдельных индивидов обуславливается их социальным контекстом и

поведением под-групп, к которым они принадлежат. SNA дает возможность определить специфику распространения различных эффектов от одних элементов сети к другим. В рамках SNA данные, как правило, представляются в виде разреженных матриц сопряженности.

В целях анализа связи между отдельными абонентами может отображаться частота звонков, их средняя длительность и другие метрики. Однако, чаще всего, веса ребер социального графа, отображающие степень тесноты связи между абонентами, представляются в виде линейной комбинации различных метрик, что позволяет одновременно учитывать множество факторов, влияющих на характер взаимодействия между узлами сети.

Наиболее широкое распространение SNA-подходы получили для решения задач прогнозирования оттока клиентов. На базе мульти-реляционных социальных сетей извлекаются различные метрики, позволяющие выявить абонентов, склонных к оттоку и произвести целевое действие для его предотвращения [42].

Помимо прогнозирования оттока, SNA применяется телекоммуникационными компаниями для решения следующих задач:

- мониторинга поведения абонентской базы в режиме реального времени;
- выявления степени лояльности определенных сегментов абонентской базы;
- построения стратегии коммуникаций;
- оптимизации взаимодействия с клиентами и рекламного таргетинга;

Для достижения синергетического эффекта при построении аналитики, телекоммуникационные компании применяют данные различных типов, так как описанные выше типы данных по-отдельности имеют значительно меньшую ценность, чем их комбинация. Интегрируя различные данные при помощи методологий CRISP-DM и ASUM-DM, возникает возможность плодотворного экспериментирования и выявления новых закономерностей в данных для извлечения ценности.

Накопленные данные различных типов используются для формирования аналитической информации, необходимой для принятия решений в определенном бизнес-контексте, в связи с этим особое внимание требует формализация задач анализа поведения потребителей в различных бизнес-контекстах.

2.3 Задачи анализа поведения потребителей в сфере телекоммуникаций

Накопление исторических данных телекоммуникационными компаниями, а также эволюция аналитической инфраструктуры от систем

отчетности до систем углубленной аналитики с применением инструментария машинного обучения привело к масштабному внедрению аналитических подходов.

Согласно отчету компании Ernst & Young, телекоммуникационные компании выделяют внедрение аналитических решений на основании технологии Big Data как приоритетное направление модернизации операционной модели [9]. Помимо потенциала внедрения аналитических решений, в отчете проблема приватности данных и регуляторные аспекты выделяются как ключевые ограничительные факторы развития бизнес-моделей, построенных на данных [9]. В отчете также выделяются стратегические приоритеты, обозначенные телекоммуникационными компаниями-респондентами. Среди них наиболее актуальными являются следующие: управление клиентским опытом, управление издержками и эффективностью, модернизация сети, повышение операционной эффективности бизнеса, улучшение ИТ-систем и бизнес-процессов. Каждый из выделенных приоритетов является отдельным бизнес-контекстом, который может быть формализован в виде аналитической задачи для последующей реализации.

Цели реализации аналитических проектов должны находиться в строгом соответствии со стратегическими целями компании. Приведение таких целей к соответствию порождает необходимость разработки стратегии данных (от англ. Data Strategy) [43]. Помимо компонентов администрирования и управления данными, важным остается определение роли данных в бизнес-процессах. Несмотря на большой потенциал совершенствования бизнес-процессов при реализации аналитических проектов, общий вектор развития компании главенствует над целями аналитики и определяет их. В связи с этим, первостепенная задача, решаемая телекоммуникационными компаниями при внедрении аналитических решений, заключается в формализации бизнес-потребностей и представлении их в виде аналитической задачи. Далее такое представление используется в аналитическом цикле CRISP-DM для выработки решения.

Для формализации определяются класс аналитической задачи, входные данные, целевой показатель, применяемые алгоритмы и способы реализации проекта, а также четкие бизнес-цели.

На основании проведенного анализа представляется целесообразным выделить следующие бизнес-контексты в качестве точек приложения аналитики больших данных [44]:

- Прогнозирование оттока клиентов;
- Таргетинговая реклама продуктов;
- Предсказание нагрузки на сеть;
- Клиентское обслуживание: речевая аналитика;
- Клиентское обслуживание: чат-боты;
- Сегментация клиентов;
- Предсказание жизненной ценности клиентов;

- Анализ жизненного цикла продуктов;
- Предписывающий анализ сервисов;
- Распознавание мошеннических схем;

Ценность реализации аналитического проекта для бизнеса и экономический эффект от внедрения зависят от качества бизнес-анализа, проводимого в целях выявления потребностей и обоснования решений, описывающих возможные пути реализации изменений.

В связи с этим актуальным является построение формализованных представлений аналитических задач, соответствующих целям бизнеса, а также рассмотрение опыта внедрения телекоммуникационными компаниями.

Источники данных представляются в соответствии с сокращениями, представленными ниже:

- Данные о клиентах – SD;
- Данные об использовании услуг – CDR;
- Гео-данные – GEO;
- Информация из каналов заботы о клиенте – CC;
- Данные о взаимодействии клиентов – SNA;

Прогнозирование оттока клиентов

Цель. Снижение оттока клиентов при одновременном снижении издержек на мероприятия по удержанию;

Задача. Построение предиктивной модели скоринга для выявления клиентов, склонных к оттоку;

Целевой показатель. Вероятность оттока на горизонте прогнозирования модели;

Описание процедуры. Фазы 1-6 цикла CRISP-DM. Формирование выборки, формирование предикторов, отбор предикторов модели, тестирование модели, оценка качества [45];

Входные данные. Для формирования предикторов модели используются следующие типы данных: SD, CDR, GEO, CC, SNA;

Формализация. Задача бинарной классификации;

Алгоритмы и методики. XGBoost, логистическая регрессия, алгоритм случайного леса;

Результаты. Снижение оттока посредством своевременного взаимодействия с клиентами, склонными к оттоку, выявленных на основании модели оттока. Повышение операционной прибыли за счет увеличения показателей удержания.

Таргетинговая реклама продуктов

Цель. Предложение продуктов в соответствии с интересами клиентов и их поведенческим профилем;

Задача. Построение предиктивной модели скоринга для определения склонности приобретения определенного продукта компании;

Целевой показатель. Вероятность приобретения продукта на горизонте прогнозирования модели;

Описание процедуры. Фазы 1-6 цикла CRISP-DM. Формирование выборки, формирование предикторов, отбор предикторов модели, тестирование модели, оценка качества [46];

Входные данные. Для формирования предикторов модели используются следующие типы данных: SD, CDR, GEO, SNA;

Формализация. Задача бинарной классификации;

Алгоритмы и методики. XGBoost, логистическая регрессия, алгоритм случайного леса;

Предсказание нагрузки на сеть

Цель. Прогнозирование нагрузки на сеть в целях управления сетевой инфраструктурой;

Задача. Построение имитационной модели для оценки нагрузки на определенные узлы сетевой инфраструктуры в целях планирования сети;

Целевой показатель. Регрессионная оценка нагрузки на сеть на узлах сетевой инфраструктуры;

Описание процедуры. Фазы 1-6 цикла CRISP-DM. Формирование выборки, формирование предикторов, отбор предикторов модели, тестирование модели, оценка качества [47];

Входные данные. Для формирования предикторов модели используются следующие типы данных: SD, CDR, GEO, CC, SNA;

Формализация. Задача регрессии;

Алгоритмы и методики. XGBoost, логистическая регрессия, алгоритм случайного леса;

Клиентское обслуживание: речевая аналитика

Цель. Внедрить систему речевой аналитики для оптимизации работы call-центра компании;

Задача. Построение системы оценки качества обслуживания клиентов, обратившихся в call-центр компании;

Целевой показатель. Сводный показатель качества обслуживания клиентов, оцененный на основании суб-метрик;

Описание процедуры. Транскрибация записи разговора в текст, аналитика расшифровки записи при помощи алгоритмов обработки текста (методы NLP), построение метрик качества работы оператора, выделение тем, именованных сущностей, а также тональности разговора, определения соответствия работы оператора регламенту обработки обращений клиентов, принятых в компании [48];

Входные данные. Для формирования предикторов модели используются следующие типы данных: SD, CDR, CC. Ключевым типом данных является аудиозапись обращения;

Формализация. Система речевой аналитики на основании алгоритмов speech2text и NLP-алгоритмов;

Алгоритмы и методики. Алгоритмы обработки естественного языка – NLP на основании подходов глубокого обучения нейронных сетей.

Клиентское обслуживание: чат-боты

Цель. Внедрить автоматическую Q&A систему взаимодействия с клиентом в цифровые каналы;

Задача. Построение чат-бота для самообслуживания клиентов;

Целевой показатель. Качество ответов чат-бота на основании обратной связи о качестве обслуживания;

Описание процедуры. На основании исторической выборки письменных обращения клиентов в цифровые каналы обучается нейронная сеть для генерации ответов, отвечающих потребностям клиентов [49];

Входные данные. В основе нейронной сети чат-бота лежит обученная на большом корпусе языковая модель;

Формализация. Генеративная нейронная сеть;

Алгоритмы и методики. Алгоритмы обработки естественного языка – NLP на основании подходов глубокого обучения нейронных сетей.

Сегментация клиентов

Цель. Построить сегментацию клиентов в целях оптимизации маркетинговой стратегии компании;

Задача. Проведение кластерного анализа для выделения сегментов клиентов на основании их поведенческих характеристик;

Целевой показатель. Эффективность маркетинговых подходов при работе с клиентами;

Описание процедуры. Формируется набор предикторов для кластерного анализа, далее применяется алгоритм кластеризации, позволяющий выделить гомогенные группы [50];

Входные данные. Для формирования предикторов модели кластеризации используются следующие типы данных: SD, CDR, GEO, CC, SNA;

Формализация. Клиентам присваиваются метки классов, соответствующие их поведенческому профилю;

Алгоритмы и методики. K-Means, C-Means, DBSCAN и др;

Предсказание жизненной ценности клиентов

Цель. Оценка жизненной ценности клиентов (от англ. Life Time Value или LTV) в целях оптимизации работы с абонентской базой;

Задача. Построение предиктивной модели оценки ценности клиентов в будущих периодах;

Целевой показатель. Разница между ожидаемой и фактической ценностью за период;

Описание процедуры. Фазы 1-6 цикла CRISP-DM. Формирование выборки, формирование предикторов, отбор предикторов модели, тестирование модели, оценка качества [51];

Входные данные. Для формирования предикторов модели кластеризации используются следующие типы данных: SD, CDR, GEO, CC, SNA;

Формализация. Для клиентов прогнозируется LTV на основании их поведенческих характеристик и набора предикторов;

Алгоритмы и методики. Оценка вероятности выживания на основании регрессии Кокса;

Анализ жизненного цикла продуктов

Цель. Проанализировать жизненный цикл продуктов, предоставляемых клиентам;

Задача. Построение описательной статистики жизненного цикла продуктов;

Целевой показатель. Отсутствует;

Описание процедуры. Фазы 1-3 цикла CRISP-DM. Оценка динамики развития продуктов [52];

Входные данные. Для формирования описательной статистики используются следующие типы данных: SD, CDR;

Формализация. Для клиентов прогнозируется LTV на основании их поведенческих характеристик и набора предикторов;

Алгоритмы и методики. Когортный анализ, анализ временных рядов;

Предписывающий анализ сервисов

Цель. Предложить рекомендации пользователям на основании их предпочтений;

Задача. Построение рекомендательной системы контента или услуг;

Целевой показатель. Качество рекомендаций, оцененное на основании оценки пользователей, касательно предлагаемых результатов;

Описание процедуры. Фазы 1-6 цикла CRISP-DM. Собираются предикторы, характеризующие предлагаемые услуги, а также пользователей, строится и тестируется рекомендательная система [53];

Входные данные. Для формирования описательной статистики используются следующие типы данных: SD, CDR, GEO, CC, SNA. Также используется информация о предпочтениях клиентов, исторические данные использования сервиса и информация из открытых источников;

Формализация. Механизм рекомендаций, построенный на исторических данных;

Алгоритмы и методики. Рекомендательные системы на основании нейронных сетей, алгоритм коллаборативной фильтрации и рекомендательная система на основании контента;

Распознавание мошеннических схем

Цель. Распознавание мошеннических схем клиентов телекоммуникационного оператора или клиентов заказчика;

Задача. Построение алгоритма детекции мошеннических схем на основании анализа поведенческого профиля;

Целевой показатель. Оценка вероятности мошенничества;

Описание процедуры. Фазы 1-6 цикла CRISP-DM. Собираются данные и формируется набор предикторов, на основании предикторов и исторических данных анализируются поведенческие аномалии, указывающие на мошенническое поведение клиента [54];

Входные данные. Для формирования описательной статистики используются следующие типы данных: SD, CDR, GEO, CC, SNA;

Формализация. Система детекции поведенческих паттернов, указывающих на мошенничество;

Алгоритмы и методики. Алгоритмы обучения с учителем и без учителя;

Выводы по главе 2

В результате анализа подходов к построению аналитических экосистем, в контексте сфере телекоммуникаций могут быть сделаны следующие выводы:

Важнейшим условием эффективного управления экосистемой данных телекоммуникационной компании является оптимальный выбор абстрактной схемы формализации процессов управления данными. В результате сравнительного анализа методологий управления данными, в качестве наиболее применимой была определена методология DAMA-DMBOK. При рассмотрении подходов к извлечению ценности из данных, были проанализированы 3 методологии: SEMMA, KDD и CRISP-DM. CRISP-DM соответствует современным практикам построения систем углубленной аналитики, в связи с чем, данный подход может быть рекомендован к внедрению телекоммуникационными компаниями в качестве методологической основы для извлечения ценности из данных. Помимо методологий управления данными и способами извлечения из них ценности, также были выделены организационные схемы, учитывающие цели построения высоко эффективной аналитической культуры. Принимая во внимание ориентацию на данные в процессах принятия решений на основании анализа можно выделить гибридную организационную модель как наиболее эффективную.

В результате анализа подходов управления экосистемой данных был сформирован набор методологий, при использовании которых достигается наибольшая эффективность реализации потенциала данных.

Многообразие данных, которыми обладают телекоммуникационные компании, позволяет им создавать комплексную аналитическую картину.

Однако, для реализации потенциала данных, необходимо понимание их природы. Рассмотренные типы данных имеют наибольшую ценность при построении аналитических решений в случае их комбинированного использовании. Внедрение новых подходов поведенческого анализа в бизнес-процессы является ключевым приоритетом телеком-компаний, ориентированных на извлечение ценности из данных.

Рассмотренные выше бизнес-контексты применения аналитики больших данных демонстрируют исключительную ценность аналитической экосистемы компании для реализации инновационных бизнес-моделей. Помимо анализа способов повышения операционной эффективности, также необходим анализ результатов, оказываемых внедряемыми решениями по обработке массивов больших данных [55].

В таблице 2.3 систематизированы задачи анализа данных и бизнес-результаты.

Таблица 2.3 – задачи анализа данных и бизнес-результаты.

Бизнес-контекст	Бизнес-результат
Прогнозирование оттока	Проведение мероприятий по удержанию клиентов
Таргетированная реклама	Формирование персонализированных предложений
Предсказание нагрузки на сеть	Оптимизация работы телекоммуникационной инфраструктуры
Клиентское обслуживание: речевая аналитика	Углубленная аналитика систем клиентского обслуживания
Клиентское обслуживание: чат-боты	Система автоматической генерации ответов на клиентские запросы
Сегментация клиентов	Адаптация стратегии взаимодействия с различными сегментами
Предсказание жизненной ценности клиентов	Анализ динамики абонентской базы на основании показателей жизненной ценности
Анализ жизненного цикла продуктов	Повышение эффективности управления жизненным циклом продуктов
Предписывающий анализ сервисов	Формирование персонализированных рекомендаций медиа-контента
Распознавание мошеннических схем	Внутренняя и внешняя fraud-аналитика

Примечание – Разработка автора

ГЛАВА 3

ВНЕДРЕНИЕ ПРОЦЕДУРЫ СКОРИНГА КЛИЕНТОВ МЕДИА-СЕРВИСА ДЛЯ КОМПАНИИ velcom | A1

3.1 Анализ эффективности процедуры скоринга медиа-сервиса и бизнес-анализ

Современные телекоммуникационные компании применяют множество инновационных конкурентных моделей, позволяющих диверсифицировать источники дохода. Одной из наиболее прогрессивных является контентная бизнес-модель, что в полной мере отражает текущие тренды в отрасли: конвергенция индустрии телекоммуникаций с медиа, повышение роли контента среди предоставляемых услуг, а также создание телеком-компаниями контентных платформ.

Среди контентных бизнес-моделей SVoD (от англ. Subscription Video on Demand или предоставление видео по подписке) бизнес-модель является наиболее распространенной. Модель SVoD включает возможность доступа в течение определенного периода к контенту в рамках каталога OTT-провайдера. Ценность SVoD бизнес-модели для компании выражается в следующем:

- Видео-контент является важным элементом конвергентной бизнес-стратегии;
- Потребление видео-контента клиентами приводит к повышению лояльности к телеком-оператору и более глубокому погружению в продуктовую экосистему компании;
- Видео-контент является одним из наиболее «тяжелых» с точки зрения потребляемого трафика типом контента;
- Также SVoD бизнес-модель в полной мере удовлетворяет целям телеком-компаний по диверсификации источников дохода;
- SVoD является гибким механизмом монетизации;

Белорусская телекоммуникационная компания velcom | A1 успешно внедрила данную бизнес-модель и предоставляет подписки посредством медиа-сервиса.

Однако, внедрение контентной стратегии velcom | A1 влечет за собой появление новых вызовов. Рост востребованности контента приводит к его удорожанию, активизация контентной стратегии и повышение конкуренции на рынке контент-платформ обуславливают необходимость предоставления клиентам уникального медиа-контента высокого качества. Вышеперечисленные факторы обуславливают необходимость в планировании развития медиа-сервиса.

В ходе бизнес-анализа процессов управления медиа-сервисом были выявлены ключевые направления его развития:

- Управление программами лояльности;

- Таргетинговая реклама контента и оценка бюджета интернет-рекламы;
- Оценка ожидаемой выручки сервиса;
- Сегментация клиентов на основе их поведенческих профилей.

Сегментация по признаку ценности клиентов была обозначена в качестве приоритетной задачи, удовлетворяющей выделенным направлениям развития медиа-сервиса.

3.1.1 Описание текущей процедуры скоринга

Для построения скоринга на основании текущей процедуры рассматриваются клиенты, имевшие списания за подписки в последний месяц или находящиеся в статусе «деактивированный», «приостановленный» в течение не более одного месяца.

Расчет скоринга производится по следующей формуле:

$$\text{Score} = w1*\text{Score1} + w2*\text{Score2} + w3*\text{Score3}$$

Где:

- Score1 – присуждаем 10% клиентов с высоким средним чеком клиента компании velcom | A1 - 3 балла, 70% остальных - 2 балла, оставшимся - 1 балл;
- Score2 – аналогичная процедура для списаний;
- Score3 - всем клиентам, которые генерируют мобильный интернет-трафик последние 3 месяца - 3 балла, всем клиентам, которые не генерируют трафик в последнем месяце - 1 балл, оставшимся - 2 балла;
- w1, w2, w3 – эвристические весовые коэффициенты (w1 = 0.3, w2 = 0.5, w3 = 0.2.).

Скоринг используется для разбиения абонентской базы на 3 сегмента. В зависимости от скоринга, клиенты могут быть отнесены к одному из трех сегментов:

- 10% верхних перцентилей - сегмент «лучшие»;
- 70% средних перцентилей – сегмент «рентабельные»;
- 20% нижних перцентилей – сегмент «убыточные».

3.1.2 Обоснование необходимости реинжиниринга бизнес-процессов скоринга

Существующая процедура скоринга абонентской базы обладает рядом недостатков, препятствующих эффективной реализации контентной бизнес-стратегии компании velcom | A1. Среди них можно выделить следующие:

- Применение эвристических весовых факторов для выставления скоринга;
- Учет только точечных факторов
 - Учитывается выручка только за последний месяц;

- Не учитывается кумулятивный срок жизни клиентов медиа-сервиса;
- Не учитывается ожидаемая ценность клиентов (expected value);
- Ручная процедура расчета на ежемесячной основе;
- Низкая стабильность скоринга месяц-к-месяцу на индивидуальном уровне;
- Учет ограниченного количества факторов, влияющих на ожидаемую выручку.

На графике ниже представлено распределение клиентов медиа-сервиса по децильным группам существующего скоринга за один отчетный месяц. Очевидно отсутствие ординальной зависимости скоринга и фактической выручки в каждой группе. Так децили №6 и №7 выбиваются и нарушают бизнес-логику скоринга: более низкому скорингу соответствует более высокая средняя фактическая выручка (Рисунок 3.1).

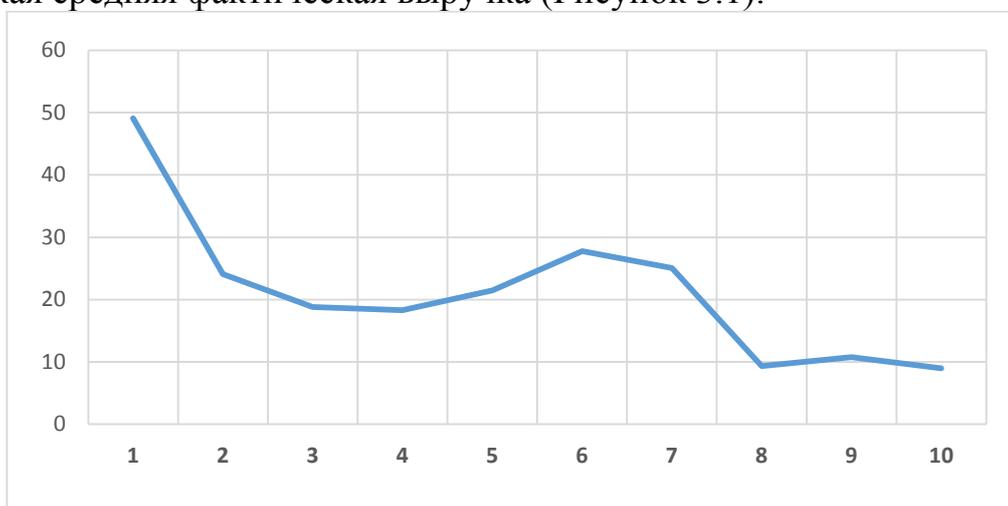


Рисунок 3.1 – Средняя выручка клиентов в децильных группах текущей процедуры скоринга

Примечание – Разработка автора

Также рассматривалась форма децильного распределения существующего скоринга за 2 месяца (Рисунок 3.2). На агрегированном уровне скоринг имеет стабильный вид, однако, на индивидуальном уровне происходит значительное смещение от месяца к месяцу. Это также нарушает бизнес-логику, так как, например, в одном месяце клиент мог получить высокий скоринг и быть отнесен в премиальный сегмент, а в другом – низкий и оказаться в нейтральном сегменте. Для оценки индивидуальных различий между двумя рассматриваемыми месяцами, применялись две метрики: MAD - среднее абсолютное отклонение (Mean Absolute Difference) и MAPD - среднее абсолютное процентное отклонение (Mean Absolute Percentage Difference). MAD составила 0.15, MAPD составил 9%.

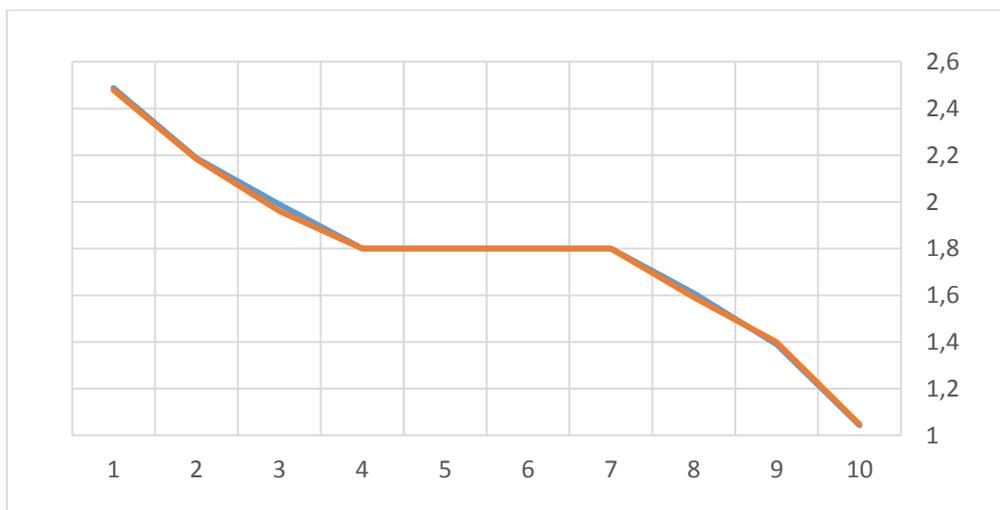


Рисунок 3.2 – Децильные распределения текущего скоринга за 2 отчетных месяца

Примечание – Разработка автора

В качестве методического подхода для реализации проекта был выбран аналитический цикл CRISP-DM.

Ниже рассмотрены шесть данной методологии.

Цели проекта. На базе приведенного обоснования необходимости реинжиниринга бизнес процессов были сформулированы следующие цели:

Построение процедуры скоринга клиентов SVoD медиа-сервиса по признаку генерируемой выручки за предыдущий и будущий периоды, внедрение и интеграция процедуры в существующую экосистему данных.

3.1.3 Критерии успешности проекта и требования к новой процедуре скоринга

На основании бизнес-анализа были выделены 4 обобщенных критерия успешности выполнения проекта по внедрению новой процедуры скоринга и их подкритерии.

1. Устранение недостатков существующей процедуры;
 - а. Отказ от эвристических весовых коэффициентов;
 - б. Использование полноты информации для формирования скоринга, увеличение количества релевантных факторов для построения скоринга;
 - с. Соблюдение бизнес-логики используемых для скоринга факторов;
2. Построение прогнозной оценки ожидаемой выручки путем построения новой интеллектуальной процедуры скоринга;
 - а. Различия ожидаемой выручки в сравнении с фактической не больше 10%;

- b. Взаимное распределение фактической ценности, генерируемой клиентами, в соответствии со скорингом должно быть ординальным;
 - c. Стабильность скоринга на индивидуальном уровне – различие не более 5%;
3. Автоматизация процедуры расчета скоринга;
- a. Последующее участие аналитиков в поддержании процедуры скоринга не более двух часов на итерацию обновления модели;
 - b. Построение автоматической системы выявления деградации модели: разница между ожидаемой и фактической выручкой превышает 7%;
 - c. Автоматический перерасчет предикторов на основании периодически обновляемой аналитической витрины данных;
4. Интеграция скоринга в бизнес-системы и процесс принятия решений;
- a. Интеграция скоринга в аналитические витрины, OLAP-кубы и системы контент-платформы;
 - b. Автоматическая генерация отчетов на ежемесячной основе о динамике клиентской базы;

3.1.4 Риски реализации проекта

Также на основании методики Боэма «Software Risk Management» были выявлены 6 групп рисков реализации проекта и внедрены механизмы их предотвращения [56]. Для выявленных рисков использовалась экспертная трёхбалльная шкала (где 1 – низкая вероятность наступления / степень влияния, 3 - высокая). Была определена вероятность наступления рисков и их степень влияния, а также стратегии предотвращения рисков. Результаты анализа рисков представлены в таблице ниже:

Таблица 3.1 – Анализ рисков реализации проекта по внедрению новой процедуры скоринга

Описание риска	Вероятность наступления	Степень влияния	Совокупная оценка	Меры предотвращения рисков
Высокая стоимость реализации проекта	2	2	4	Тщательное планирование этапов реализации проекта
Нереалистичные сроки	2	1	2	Введение прогрессивной системы управления сроками реализации
Неэффективное управление требованиями и качеством	1	2	2	Контроль качества, проведение data-profiling и предоставление отчетности по мере реализации этапов цикла CRISP-DM
Низкое качество входных данных	3	3	9	Проведение мероприятий по контролю качества и тесное взаимодействие с BI-подразделением
Несоответствие аналитического решения бизнес-целям	2	3	6	Назначение установочных встреч с руководителями направления
Низкая	1	3	3	Предоставление отчетности о результатах

вовлеченность менеджмента компании				реализации проекта, а также предоставление экономического обоснования проекта
------------------------------------	--	--	--	---

Примечание – Разработка автора

Таким образом, среди выявленных рисков, наиболее значимыми оказались 3, получившие наибольшую совокупную оценку степени риска: низкое качество данных, несоответствие аналитического решения бизнес-целям и высокая стоимость реализации проекта.

3.1.5 Предварительный анализ

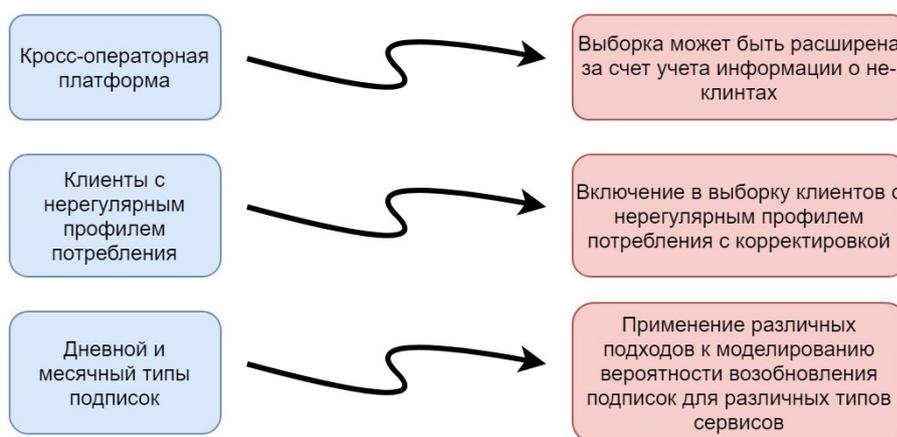
В ходе предварительного анализа на основании выделенных недостатков были сформулированы и протестированы гипотезы по улучшению существующей процедуры скоринга. Также был выявлен потенциал по оптимизации бизнес-процессов медиа-сервиса.

В целях оптимизации процедуры существующего скоринга был проведен анализ наиболее распространенных подходов к оценке ценности клиентов. За основу скоринга была взята предикторная модель регрессии Кокса, на основании которой проводилась оценка вероятности продления подписок клиентами медиа-сервисов. Более детально данная модель и ее примеры применения будут рассмотрены ниже.

При выработке бизнес-гипотез во внимание были приняты следующие аспекты и на их основании выявлен дополнительный аналитический потенциал данных:

- Рассматриваемый медиа-сервис является кросс-операторным, то есть его клиентами могут являться как клиенты мобильного оператора velcom | A1, так и клиенты других мобильных операторов;
- Для анализа может быть использована информация о клиентах с нерегулярным профилем оплаты услуг; Вероятность реактивации клиентов в рамках подписки после одного месяца перерыва составляет порядка 80%, после двух – порядка 12%, после трех – порядка 5%;
- В SVoD сервисе существует два типа подписок на контент – на месяц (Monthly) и на один день (One Day). В связи с отличием моделей жизненного цикла Monthly- и OneDay-подписок, необходима диверсификация стратегий работы с различными типами;

Для наиболее полного учета всей доступной информации на основании выявленного потенциала данных, были предложены следующие меры,



отраженные ниже (Рисунок 3.3).

Рисунок 3.3 – Анализ потенциала данных

Примечание – Разработка автора

Оценка ценности проекта и его влияние на бизнес-процессы. На первом этапе CRISP-DM была определена ценность проекта для компании в случае его успешной реализации. На основании результатов скоринга, удовлетворяющего обозначенным требованиям, управление развитием медиа-сервиса станет более эффективным за счет повышения точности прогнозирования денежного потока. Также сегментация клиентской базы позволяет реализовывать различные маркетинговые программы, направленные на повышение лояльности к экосистеме компании. Скоринг клиентов может также выступать в роли фактора таргетинговой рекламы, что позволит оптимизировать расходы на рекламу за счет формирования персональных предложений. Проведенный анализ и построенная модель в рамках проекта по оптимизации процедуры построения скоринга являются базисом для будущих аналитических проектов, связанных с медиа-сервисом.

Новая процедура скоринга была отражена в ВРМ-системе в качестве транзакций изменения и добавления бизнес-процессов.

В результате внедрения новой процедуры скоринга были изменены следующие бизнес-процессы:

- Построение периодической отчетности по медиа-сервисам;
- Процедура построения скоринга;
- Процедура взаимодействия бизнес-пользователей с результатами скоринга и сегментации;

Также были внедрены следующие бизнес-процессы:

- Мониторинг качества модели и определение степени деградации модели;
- Процедура обновления витрины данных;

Формализация задачи. В целях сегментации были определены такие свойства скоринга, которые бы удовлетворяли бизнес-логике и контексту задачи повышения эффективности управления развитием медиа-сервиса в рамках контентной бизнес-модели. Эти свойства отражены ниже:

- Учет в модели скоринга как ожидаемой выручки, так и выручки за предыдущие месяцы;
- Введение механизма премирования клиентов за лояльность и устойчивость использования медиа-сервиса;
- Использование прогностической составляющей;

На основании предварительного анализа была произведена формализация задачи построения скоринга, то есть переход от бизнес-формулировки к детализированному техническому описанию процедуры построения скоринга, учитывающего различные аспекты управления медиа-сервисом.

Для построения скоринга используются данные о клиентах, а также данные об использовании услуг в рамках медиа-сервиса за 12 месяцев. Размер выборки составляет порядка 50 тысяч клиентов в месяц. Количество подписок составляет порядка 15.

Ввиду высокой динамики базы, выражающейся в интенсивном подключении и отключении подписок клиентами, необходима разработка модели для прогнозирования ценности клиентов на временном горизонте в 3 месяца. Также исторические данные в 12 месяцев ограничивают возможности прогнозирования на более длительных временных горизонтах.

Также был сформирован план построения аналитического решения, который соотносится с методологией CRISP-DM.

На первом этапе аналитического цикла CRISP-DM производилась декомпозиция задачи. Были определены критерии успешности выполнения проекта, проведен предварительный анализ, произведена оценка рисков и был построен план реализации проекта, а также сформулированы требования к новой процедуре скоринга.

На втором этапе производится анализ данных.

На третьем этапе данные подготавливаются для последующего их использования для построения моделей.

На четвертом этапе моделирования происходит отбор наиболее значимых предикторов и строится модель, а также оценивается ее качество.

На пятом этапе оценивается общее качество модели и принимается решение о ее внедрении.

На шестом этапе происходит интеграция решения и автоматизация бизнес-процессов.

3.2 Оптимизация процедуры скоринга клиентов медиа-сервиса

Для прогнозирования генерируемой выручки существует ряд общепринятых подходов. Среди них наиболее популярными являются два из них - прогнозирование выручки на основании RFM-анализа, а также на основании регрессии Кокса.

В маркетинге методика RFM-анализа применяется для прогнозирования поведения клиента на основе его прошлых действий. RFM – аббревиатура англ. слов Recency, Frequency, Monetary, что в переводе означает новизна, частота, покупки. **Recency** – это время, прошедшее с последней активности клиента, будь то покупка, переход по ссылке или открытие письма. **Frequency** – количество действий, совершенных клиентом. **Monetary** – кумулятивные денежные затраты клиента на товары и услуги компании за период наблюдения. Однако, RFM не позволяет выявлять причины, оказывающие влияние на вероятность выживания в будущих периодах, а также не позволяет прогнозировать ожидаемую выручку.

3.2.1 Регрессия Кокса

Конвенциональные статистические методы, такие как логистическая регрессия или алгоритм случайного леса, позволяют успешно предсказывать вероятность оттока или вероятность выживания клиентов. Однако, эти модели непригодны для прогнозирования того, когда именно наступит целевое событие или как долго клиент будет пользоваться услугами компании. В связи с этим был разработан инструментарий для работы с данными о выживании (от англ. survival data).

Подходы прогнозирования ценности клиентов на основании анализа выживаемости (от англ. Survival Analysis) получили широкое распространение благодаря возможности прогнозирования времени до наступления события, а также использования цензурированных данных. Анализ выживаемости – это класс статистических моделей, позволяющих оценить вероятность наступления события. Данный тип моделей получил свое название в связи с их массовым использованием в биомедицинских исследованиях для оценки продолжительности жизни при изучении эффективности методов лечения. Позднее модели выживаемости стали применяться в страховой сфере, и социальных науках. Например, в экономике модели выживаемости используются для оценки времени, которое индивид проведет в статусе безработного до того, как найдет работу [57]. Развитие инструментария анализа выживаемости связано с именем Дэвида Кокса, автора модели пропорциональных рисков или, как ее часто называют, модели регрессии Кокса [58]. Статья Дэвида Кокса «Regression Models and Life-Tables» является одной из наиболее цитируемых статей по прикладной статистике – около 8500 цитирований.

Цензурированные данные – это либо наблюдения, которые выпали из исследования естественным образом, либо наблюдения, о которых достоверно известно, что целевое событие для них не наступило.

При проведении анализа выживаемости различают цензурирование слева и справа. Цензурирование слева проводится, если неизвестно время начала эксперимента. Например, в медицинских исследованиях часто требуется цензурирование слева (если пациент поступил в госпиталь и известно, что он выжил в течение определенного времени, однако отсутствует информация о времени начала симптомов). Цензурирование справа имеет место, если известно, когда начался эксперимент и что он закончится в точке, расположенной справа от точки начала эксперимента.

На графике ниже представлен механизм цензурирования слева и справа (Рисунок 3.4).

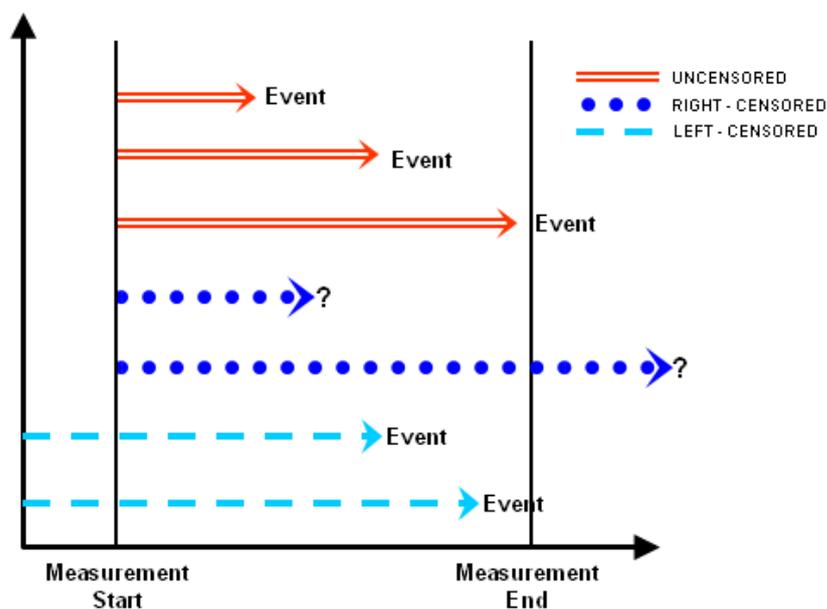


Рисунок 3.4 – Процедуры цензурирования слева и справа

Примечание – Разработка автора

При анализе данных дожития две функции представляют фундаментальный интерес – это функция дожития (от англ. survival function) и функция риска (от англ. hazard function).

Пусть переменная T является временем дожития (от англ. survival time). Эта переменная отражает время, прошедшее от начала исследования (например, время от постановки диагноза) до события (например, смерти пациента по причине этого заболевания). Значения T можно рассматривать, как имеющие распределение вероятностей (от англ. probability distribution). Предположим, что функция плотности вероятности (от англ. probability density function) случайной величины T задается как $f(T)$. Тогда распределение вероятностей функции T в момент времени t выражается формулой [59]:

Функция дожития, $S(T)$ – это вероятность, что объект исследования выживет (событие не наступит) по прошествии времени T . Тогда формула будет иметь вид:

Функция риска – это вероятность, что событие наступит в течение ничтожно малого временного интервала, учитывая, что объект доживает до

начала этого интервала. Математическим выражением для этой функции будет следующим:

$$\frac{dF(t)}{dt} = -\lambda(t) S(t)$$

Таким образом, функция риска является производной от функции кумулятивного риска. Отношение функции кумулятивного риска к функции кумулятивного дожития выражается формулами:

Насколько мы можем судить по этим формулам, функция плотности вероятности, функция риска и функция дожития – математически взаимосвязаны [60].

Отношение между риском наступления события и независимыми переменными выражается формулой:

, где:

- T – время;
- X_1, \dots, X_k – независимые переменные (предикторы).
- $\lambda_0(t)$ – основной риск наступления события (от англ. baseline hazard rate), одинаковый для всех объектов наблюдения (при условии, что значения всех независимых переменных равны 0, то есть они никоим образом не влияют на исход);
- β_1, \dots, β_k – регрессионные коэффициенты;

В этой модели регрессионные коэффициенты указывают на влияние каждого предиктора на функцию риска, и при увеличении значения предиктора на единицу, если значения остальных переменных неизменны, риск наступления события возрастает в $\exp(\beta)$ раз [58].

Таким образом, в регрессии Кокса нас будут интересовать три вида показателей: исход, период наблюдения, предикторы. Если бы нас интересовали только исход и предикторы, то мы могли бы воспользоваться для анализа методом логистической регрессии. Если бы мы были заинтересованы в оценке исхода относительно времени, то можно было бы провести анализ методом Каплана-Мейера или построения таблиц дожития [61]. Только в регрессии Кокса мы можем оценить влияние множества предикторов на исход с учетом периода наблюдения.

По форме модель пропорциональных рисков схожа с моделью логистической регрессии, при оценке последней используется метод максимального правдоподобия. Попытка использовать для оценки продолжительности жизни или других показателей критерий максимального

правдоподобия, привела к необходимости решать интегральные уравнения (в связи с наличием цензурированных случаев), что представляет собой достаточно трудоемкий процесс. Для оценки модели пропорциональных рисков Д.Р. Коксом для восстановления зависимостей при анализе выживаемости предложен метод частичного правдоподобия (от англ. *partial likelihood*), который позволяет оценить значения компонент вектора коэффициентов модели, доставляющие максимум так называемой частичной функции максимального правдоподобия, и строить эффективные в вычислительном отношении процедуры оценки, но по принципу идентичен методу максимального правдоподобия [58]. Данный метод так же показывает, какая часть информации осталась необъясненной после применения модели для наших фактических данных. Модель обладает меньшей предсказательной способностью, если значения этого показателя возрастают.

В регрессии Кокса (как и в логистической регрессии) нулевая гипотеза (предиктор не имеет связи с зависимой переменной, т.е. его регрессионный коэффициент значимо не отличается от нуля) проверяется при помощи критерия Wald. Если регрессионный коэффициент значимо отличается от нуля, то независимая переменная вносит значимый вклад в предсказательную способность модели [62].

Показатель $-2LL$ достаточно сложен для интерпретации, поэтому используется коэффициент $\exp(\)$, который показывает во сколько раз изменяется риск возникновения исхода, если значение предиктора изменяется на единицу.

Для применения регрессии Кокса необходимо соблюдение целого ряда критериев [63]: в основе регрессии Кокса имеется три базовых предположения относительно переменных, которые будут представлены первыми в списке. Следует также учитывать условия, которые являются значимыми для всех методов анализа дожития:

1. Все предикторы независимы. Если выявлено взаимное влияние независимых переменных, то в модель необходимо включить функцию взаимодействия этих факторов;
2. Все переменные линейно влияют на логарифм функции риска наступления события;
3. Риск наступления события для любых двух объектов в любой интервал времени пропорционален. Условие подразумевает предположение о мультипликативной зависимости между функцией интенсивности и логлинейной функцией ковариат и является главным для построения модели в регрессии Кокса без зависимых от времени переменных. Данное предположение также называют гипотезой о пропорциональности рисков;
4. Момент начала и окончания исследования (возникновения исхода или окончания периода наблюдения) или интервал наблюдения в единицах времени должны быть точно определены для каждого члена выборки;

5. Определение исхода и момент его возникновения также должны быть четко зафиксированы;
6. Цензурированные и нецензурированные наблюдения не должны отличаться по выживаемости друг от друга;
7. Методы оценки выживаемости и определения исхода одинаковы на протяжении всего исследования;
8. Условия, которые влияют на выживаемость, не меняются в ходе исследования;

Как правило, функция выживания и функция риска используются для описания статуса выживания клиентов во время срока наблюдения. На основании функции выживания определяется вероятность выживания после определенного момента времени T . Функция риска описывает риск события (например, отток клиентов) в промежуток времени за временем T , при условии, что клиент уже дожил до времени T .

Анализ выживаемости является распространенным подходом при построении LTV-скоринга клиентов и широко применяется в телекоммуникационной сфере для различных задач прогнозирования. По оценкам, годовой отток клиентов составляет порядка 30%, в то время как удержание клиентов обходится в 5-10 раз дешевле, чем привлечение новых [64]. В связи с этим инструментарий анализа выживаемости активно применяется при построении предиктивных моделей оттока, LTV-скоринга, а также для поведенческого моделирования.

Например, анализ выживаемости применяется для оценки ожидаемой выручки, генерируемой клиентом на основании ковариат (предикторов). В статье рассматривается способ построения оценки ожидаемой выручки на основании предсказанных вероятностей выживания клиента на протяжении горизонта прогнозирования модели. [64]

Для этого авторами используется следующая формула расчета ожидаемой выручки:

Где MM – месячная маржа клиента для трех последних месяцев для действующих клиентов или маржа за последний месяц наблюдений для новых клиентов. Месячная маржа рассчитывается либо на основании бухгалтерской отчетности, либо на основании регрессионных моделей. T – количество месяцев, для которых рассчитывается ожидаемая выручка. T может составлять от 12 до 36 месяцев; r – ставка дисконтирования, P_i – это вероятность выживания клиента в период i . P_i оценивается на основании модели выживания. В данном случае модель регрессии Кокса является ключевым элементом построения оценки ожидаемой выручки.

3.2.2 Описание процедуры построения регрессии Кокса для скоринга клиентов медиа-сервиса

В целях оценки вероятности для построения новой процедуры скоринга клиентов медиа-сервиса, также использовалась модель регрессии Кокса. Как было отмечено ранее, глубина накопленной истории для пар клиент-подписка в базе данных медиа-сервиса velcom | A1 составляет 12 месяцев, в связи с этим горизонт прогнозирования T составляет 3 месяца. Для горизонта прогнозирования в 3 месяца от использования ставки дисконтирования можно отказаться в целях упрощения модели.

На второй и третьей фазе CRISP-DM проводились подготовка данных и расчет предикторов:

В качестве показателя платежного поведения используется средняя выручка, сгенерированная клиентом в рамках медиа-сервиса за последние 3 месяца. При расчете показателя учитывается выручка только в месяцах, в которых клиент платил.

В качестве целевой переменной для прогнозирования вероятности выживания используется расчетный статус, представленный в бинарном виде. 1 – клиент платил за сервис, 0 – клиент не платил за сервис.

Для оценки вероятностей использована модель регрессии Кокса, которая опираясь на исторические данные за последние 12 мес. для каждой пары «клиент – подписка» вычисляет функцию выживаемости, т.е. оценку вероятности пролонгирования оплат в зависимости от уже накопленного количества оплаченных периодов и дополнительных предикторов. Для построения регрессии Кокса тестировались порядка 8 групп предикторов, 3 из них были выбраны для включения в финальную спецификацию:

- статус сервиса на конец последнего месяца (возможны три типа – «активный», «приостановленный», «деактивированный»);
- средний счета* за услуги мобильной связи velcom | A1 (сгруппированные по пяти группам);
- срок жизни абонента velcom | A1 (сгруппированные по трем группам);

* Для нивелирования инфляционных и сезонных эффектов в модели использованы не абсолютные значения ARPU, а отношение к среднему ARPU по базе за период.

В связи с тем, что модели жизненного цикла Monthly и OneDay подписок различаются, для каждого из типов подписок построены отдельные модели регрессии Кокса.

Поскольку исторические данные ограничены 12 мес. периодом, модель Кокса также имеет ограничение по горизонту прогнозирования: при накопленном сроке жизни (от англ. lifetime) в 12 мес. дальнейший прогноз не может быть получен. Для того, чтобы обойти это ограничение была построена регрессионная модель экстраполяции «кривых выживаемости». В качестве регрессоров использовались те же самые предикторы, что и в модели Кокса.

На графике ниже показана зависимость средней вероятности пролонгации платежей для месячных и однодневных подписок в зависимости от кол-ва месяцев с платежами в прошлом (Рисунок 3.5).

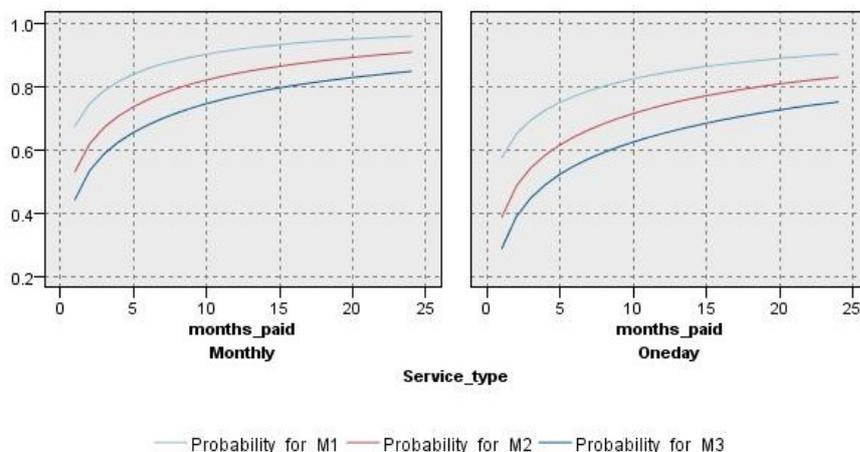
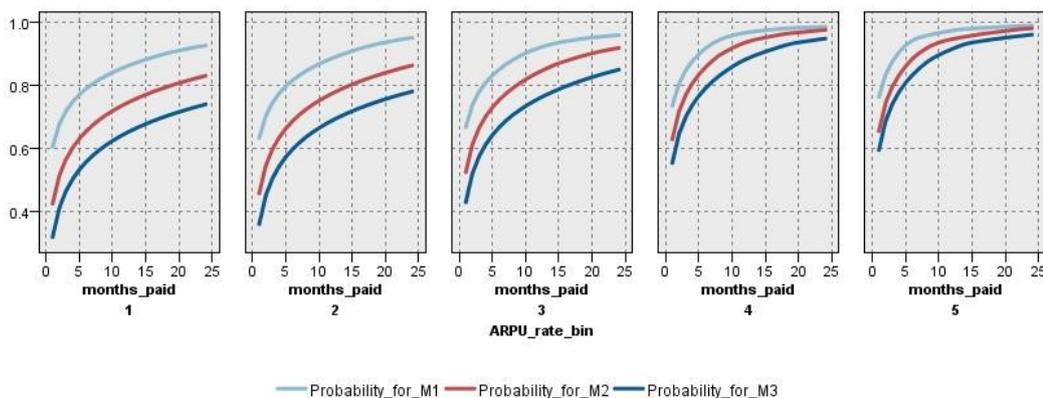


Рисунок 3.5 – Зависимость средней вероятности пролонгации платежей за подписки для типов подписок Monthly и Oneday
Примечание – Разработка автора

На графике представлены вероятности наличия будущих платежей в



зависимости от групп ARPU velcom | A1 и групп срока жизни velcom | A1.

Рисунок 3.6 – Вероятность пролонгации подписок в зависимости от принадлежности к группе ковариат ARPU velcom
Примечание – Разработка автора

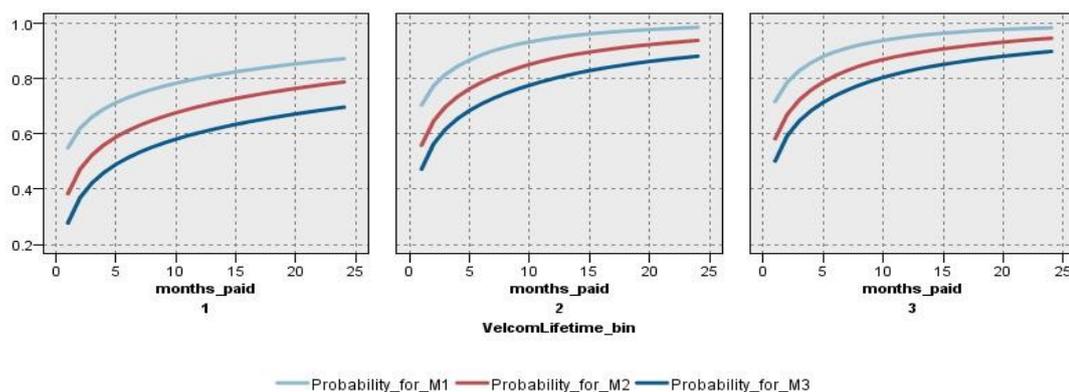


Рисунок 3.7 – Вероятность пролонгации подписок в зависимости от принадлежности к группе ковариат срока жизни в velcom
 Примечание – Разработка автора

3.2.3 Описание новой процедуры построения скоринга

На четвертом этапе цикла CRISP-DM оценивалось качество модели. Также результаты моделирования были использованы при построении скоринга клиентской базы, отвечающей требованиям, сформулированным на первом этапе CRISP-DM.

Результатом построения регрессии Кокса является возможность прогнозирования вероятности пролонгации подписок клиентами на протяжении трех последующих месяцев.

Скоринг (Score), используемый для сегментации клиентов, состоит из двух компонентов – фактической выручки (Score_P) и ожидаемой (Score_F) и имеет следующий вид:

, где:

- Score_P – скоринг на основании показателей по подписке за прошлый период (фактический);
- Score_F – скоринг на основании ожидаемой выручки по подписке в следующие 3 месяца;
- k – вес вклада выручки будущего периода, принимает значение 0.5;
- Σ – суммирование по всем платным подпискам, активированным у клиента;

, где:

- Actual Revenue – сумма списаний по медиа-сервису на клиенте за последние 3 месяца;
- –премирующий коэффициент за лояльность, который принимает значение от 0 до 0.5 в зависимости от накопленного срока жизни для конкретной подписки. Если срок жизни клиента в рамках медиа-сервиса меньше 3 месяцев, то премирующий коэффициент равен нулю ($= 0$), если срок жизни больше 3 месяцев, коэффициент рассчитывается по сигмоидальной функции активации, имеющий вид S-кривой (Рисунок 3.8).

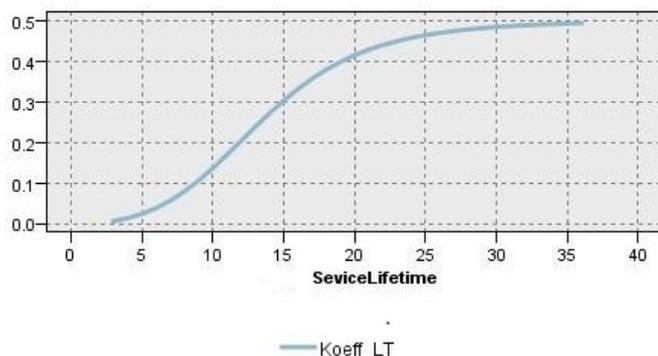


Рисунок 3.8 – Сигмоидальная функция для расчета премирующего коэффициента за срок жизни

Примечание – Разработка автора

Гистограмма распределения скоринга Score_P имеет следующий вид (Рисунок 3.9):

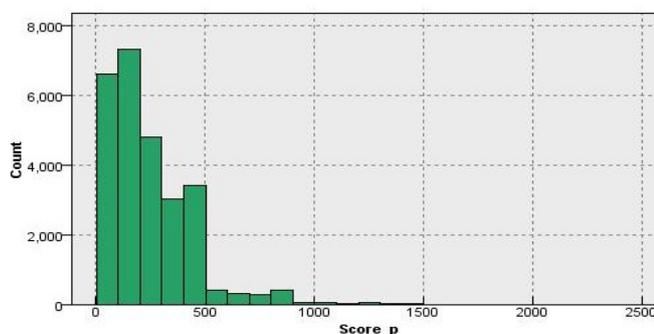


Рисунок 3.9 – Гистограмма распределения скоринга Score_P

Примечание – Разработка автора

,

, где:

- Вероятности – вероятности пролонгации подписок в месяцы $M + 1, M + 2, M + 3$;
- Revenue_Mean - средняя сумма списаний за последние 3 месяца, когда клиент платил.

Гистограмма распределения скоринга Score_F имеет следующий вид (Рисунок 3.10):

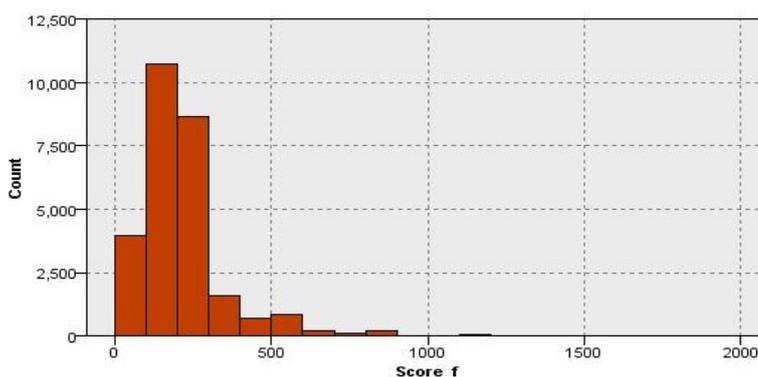


Рисунок 3.10 – Гистограмма распределения скоринга Score_F

Примечание – Разработка автора

Совокупный скоринг Score рассчитывается по формуле [1] и имеет следующий вид:

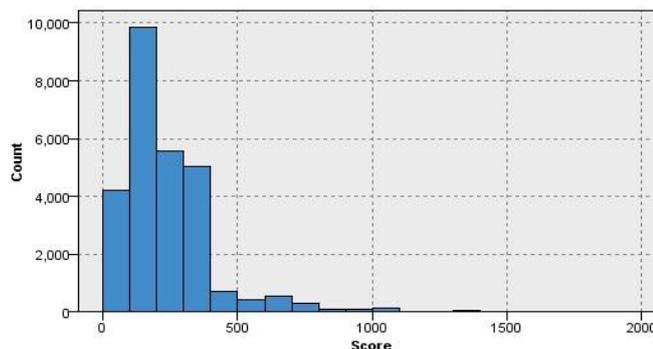


Рисунок 3.11 – Гистограмма распределения скоринга Score

Примечание – Разработка автора

Скоринг складывается из двух компонент Score_P и Score_F, показано на гистограмме (Рисунок 3.12):

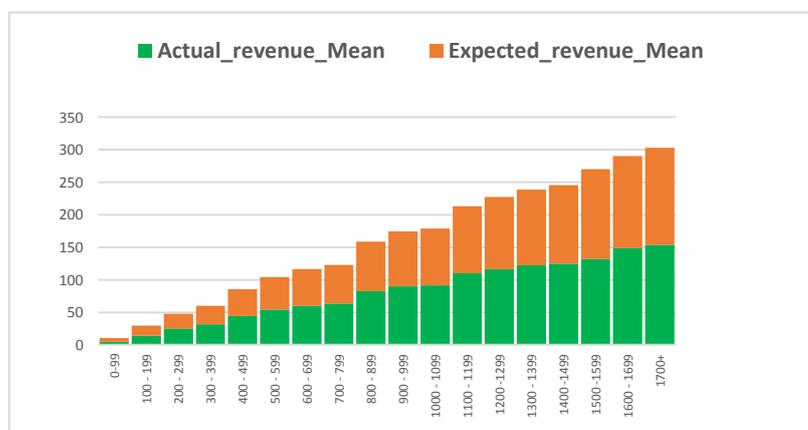


Рисунок 3.12 – Гистограмма распределения скоринга Score

Примечание – Разработка автора

Таким образом, новый скоринг учитывает, как прошлую выручку, генерируемую клиентом, так и ожидаемую за счет введения прогностической составляющей в виде оценки вероятностей пролонгации подписок при помощи модели регрессии Кокса. Также в новом скоринге учитывается срок жизни клиента и используется премирующий коэффициент за лояльность к медиа-сервису, что также соответствует целям компании повышения лояльности к экосистеме в целом. Помимо этого, за счет прогнозирования вероятностей пролонгации подписок на уровне «клиент-подписка» достигается более высокая точность оценки, так как для двух типов подписок OneDay и Monthly паттерны поведения клиентской базы значительно отличаются.

3.3 Анализ результатов внедрения новой процедуры скоринга

На пятом этапе цикла CRISP-DM была произведена оценка общего качества модели и произведено обоснование целесообразности внедрения скоринга в бизнес-процессы компании. Также была проанализирована полнота соответствия построенной процедуры скоринга заявленным на первой фазе критериям успешности реализации проекта.

На шестом этапе производилось внедрение новой процедуры скоринга в бизнес-процессы компании.

Согласно сформулированным требованиям была произведена проверка соответствия построенного скоринга обозначенным критериям качества.

На первом этапе были выделены такие недостатки процедуры скоринга, как: использование эвристических весовых коэффициентов для построения скоринга, использование ограниченного числа факторов, нестабильность процедуры скоринга на индивидуальном уровне, а также соответствие используемой процедуры скоринга требованиям бизнес-логики компании.

Все выделенные недостатки существующей модели скоринга исправлены, что означает выполнение критериев успешности в пункте 1. Результаты представлены в таблице 3.2.

Таблица 3.2 – Выполнение критериев успешности реализации проекта (пункт 1)

	Критерий успешности	Результаты нового скоринга
1. Устранение недостатков существующей процедуры;	Устранение недостатков	Все выделенные недостатки существующей процедуры устранены, критерии успешности выполнены
а. Отказ от эвристических весовых коэффициентов;	Использование эмпирических оценок	Скоринг складывается из двух компонент, построенных на основании статистических моделей

в. Использование полноты информации для формирования скоринга;	Рассмотрение большего числа релевантных факторов	Были рассмотрены порядка 8 факторов, 3 из которых применяются в финальной спецификации
с. Соблюдение бизнес-логики используемых для скоринга факторов;	Отсутствие нарушений бизнес-логики	Бизнес-логика соблюдается

Примечание – Разработка автора

Таким образом, критерии успешности реализации проекта пункта 1 выполняются.

На графике представлено распределение клиентов медиа-сервиса по децильным группам существующей и новой процедуры скоринга за один отчетный месяц (Рисунок 3.13).

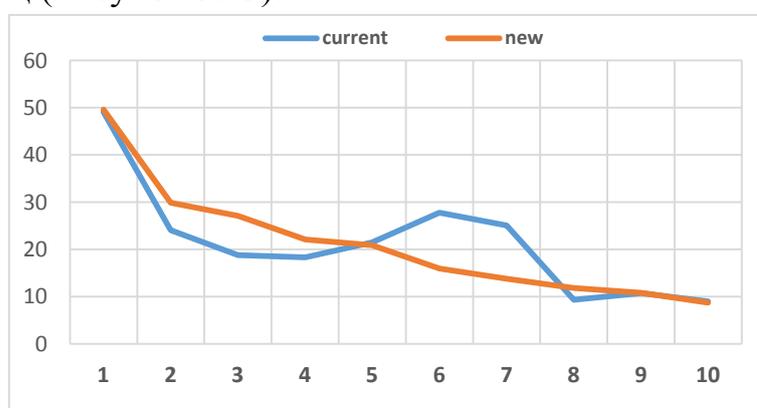


Рисунок 3.13 – Средняя выручка на абонента по децильным группам существующего и нового скоринга

Примечание – Разработка автора

В отличие от существующей процедуры, в новой процедуре скоринга присутствует ярко выраженная ординальная зависимость скоринга и фактической выручки за тот же период в каждой группе. То есть более высокому скорингу соответствует более высокая средняя фактическая выручка, что полностью соответствует бизнес-логике и позволяет на основании новой процедуры скоринга проводить сегментацию клиентов по признаку ожидаемой ценности.

Важной особенностью новой процедуры скоринга является возможность выделения наиболее платежеспособного сегмента.

Рассмотрим формы децильных распределений существующей и новой процедур скоринга за 2 отчетных месяца. Как на агрегированном, так и на индивидуальном уровне, новая процедура скоринга имеет более стабильный вид. Такое свойство скоринга в полной мере отвечает бизнес-логике: за счет использования компоненты Score_P миграция из одного сегмента, выделенного на основании скоринга, в другой, будет происходить более плавно. Это можно подтвердить при помощи сравнения метрик MAD - среднее абсолютное отклонение (Mean Absolute Difference) и MAPD - среднее абсолютное процентное отклонение значений скоринга за два месяца

по существующей и новой процедурам скоринга. Метрики представлены в таблице 3.3.

Таблица 3.3 – Выполнение критериев успешности реализации проекта

	Текущая процедура	Новая процедура
Mean Absolute Difference	0,15	0,04
Mean Absolute Percent Difference	9%	3%

Также свойство стабильности скоринга на индивидуальном уровне прослеживается на графике распределения скоринга по децильным группам за два отчетных месяца (Рисунок 3.14).

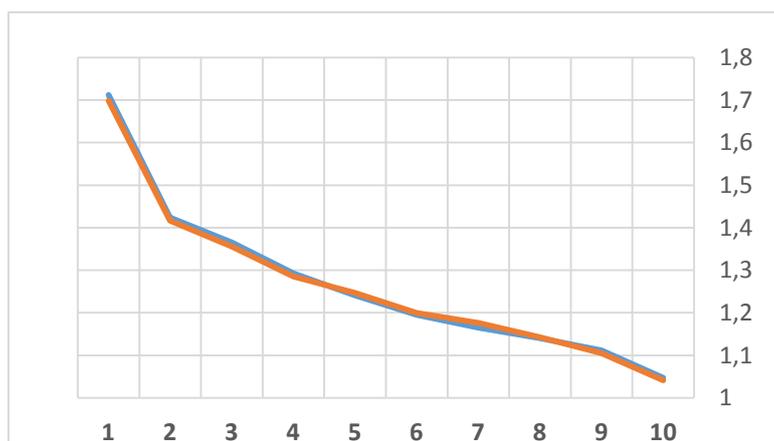


Рисунок 3.14 – Децильные распределения нового скоринга за два отчетных месяца

Примечание – Разработка автора

В результате построения оценки ожидаемой выручки Score_F на 3 последующих месяца, был проведен ретроспективный анализ, в рамках которого было оценено качество прогнозирования ожидаемой выручки. Данные сравнивались на агрегированном уровне. Сравнивалась ожидаемая выручка, построенная по формуле

и фактическая суммарная выручка за тот же период. График показывает, что на агрегированном уровне оценка ожидаемой выручки является достаточно адекватной. Общее отклонение на отложенной выборке по суммарной выручке медиа-сервиса составило 3%. Метрика MAPE, рассчитанная для децильных распределений составила 12%. Более заметно отклонение в крайних децилях – первом и десятом. Сравнение фактической и ожидаемой выручки представлено на графике (Рисунок 3.15).

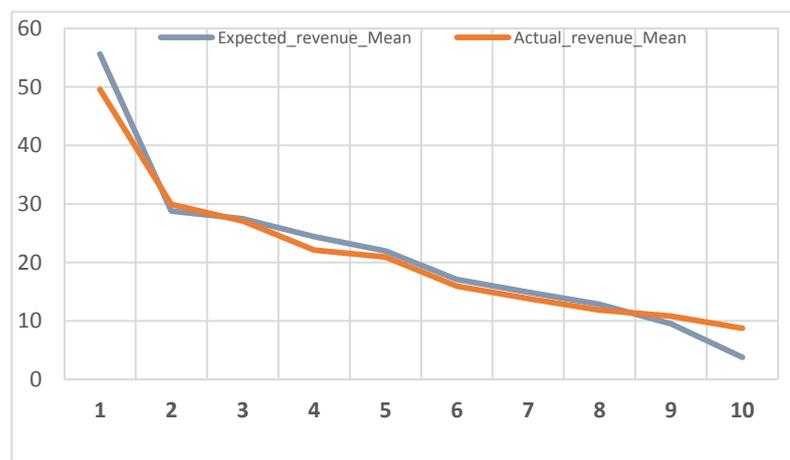


Рисунок 3.15 – Распределение ожидаемой и фактической выручки по децильному распределению нового скоринга

Примечание – Разработка автора

Возвращаясь к обозначенным на первом этапе критериям успешности проекта, можно констатировать, что все цели, выделенные в пункте 2 выполнены успешно. Результаты отражены в таблице 3.4.

Таблица 3.4 – Выполнение критериев успешности реализации проекта (пункт 2)

	Критерий успешности	Результаты нового скоринга
2. Построение прогнозной оценки ожидаемой выручки;	Построение оценки	Оценка построена, все критерии успешности выполнены
а. Различие ожидаемой выручки в сравнении с фактической;	10%	3%
б. Взаимное распределение фактической ценности в соответствии со скорингом должно быть ординальным;	Выполнение условия ординальности скоринга	Условие ординальности скоринга выполняется
с. Стабильность скоринга на индивидуальном уровне;	5%	3%

Примечание – Разработка автора

В пункте 3 критериев успешности реализации проекта была поставлена задача автоматизации процедуры скоринга, предусматривающая снижение степени участия аналитиков в последующих мероприятиях поддержки и обновления скоринговой модели, построение системы выявления деградации модели, а также построение аналитической витрины данных.

Результаты приведены в таблице 3.5.

Таблица 3.5 – Выполнение критериев успешности реализации проекта (пункт 3)

	Критерий успешности	Результаты нового скоринга
3. Автоматизация процедуры расчета скоринга;	Автоматизация процедуры расчета скоринга, предусматривающая минимизацию использования ресурсов на поддержание модели	Автоматизация выполнена успешно
а. Минимизация последующего участия аналитиков в поддержании процедуры скоринга;	Ограничение времени на обновление и поддержание модели аналитиками не должно превышать 2-часов	Фактическое участие аналитиков составляет порядка 4 часов
б. Построение автоматической системы выявления деградации модели;	Внедрение системы мониторинга уровня деградации модели	Система внедрена
с. Автоматический перерасчет предикторов на основании периодически обновляемой аналитической витрины данных;	Внедрение периодически обновляемой аналитической витрины данных	Витрина внедрена

Примечание – Разработка автора

По результатам оценки скорости проведения итерации обновления модели, участие аналитика составило 4 часа, что означает невыполнение подкритерия успешности. Таким образом один из трех подкритериев успешности не выполняется, тем не менее, критерий 3 выполняется, так как проведена автоматизация процедуры расчета нового скоринга.

В пункте 4 критериев успешности ставилась задача внедрения скоринга в бизнес-системы компании velcom | A1, что было успешно реализовано. Также был налажен процесс построения отчетности посредством написания программы генерации отчета на основании данных о медиа-сервисе.

Таким образом, выполнены все критерии успешности реализации проекта:

- Устранены недостатки существующей процедуры;
- Построена модель прогнозной оценки ожидаемой выручки;
- Произведена автоматизация построенного решения;
- Результаты внедрены в бизнес-системы компании.

Также можно говорить о соответствии полученных результатов бизнес-целям проекта. На основании построенного скоринга была произведена сегментация клиентов медиа-сервиса. На диаграмме рассеяния показано взаимное распределение ожидаемой и фактической выручки (Рисунок 3.16).

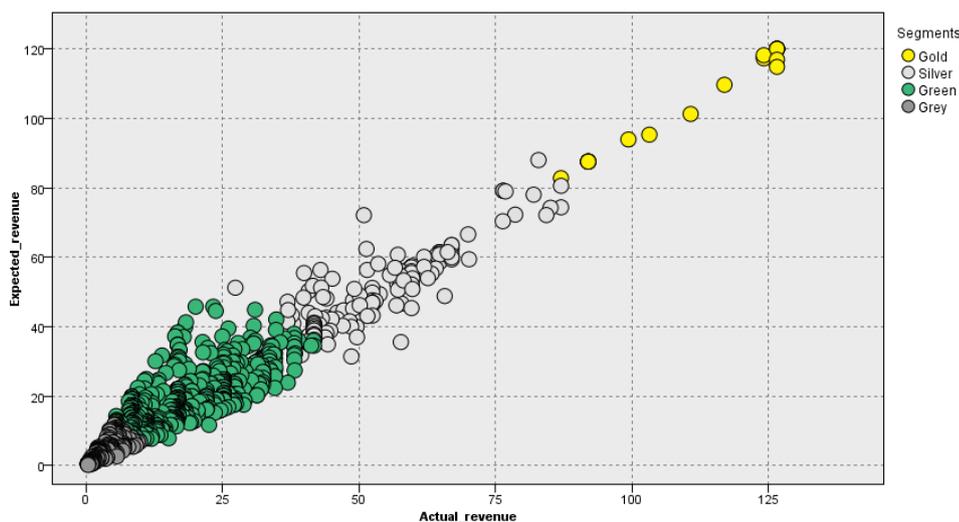


Рисунок 3.16 – Диаграмма рассеяния совместного распределения ожидаемой и фактической выручки в зависимости от сегмента

Примечание – Разработка автора

Также на основании новой процедуры скоринга были выделены 4 сегмента: Gold, Silver, Green, Grey. Распределение клиентов по сегментам было произведено на базе скоринговой оценки по принципу, отраженному в таблице 3.6.

Таблица 3.6 – Алгоритм распределения клиентов по сегментам на основании их скоринга

	Перцентили
Gold	1
Silver	2-10
Green	11-80
Grey	81-100

Примечание – Разработка автора

Соответствие распределений ожидаемой и фактической выручки по сегментам отражено на круговых диаграммах (Рисунок 3.17).

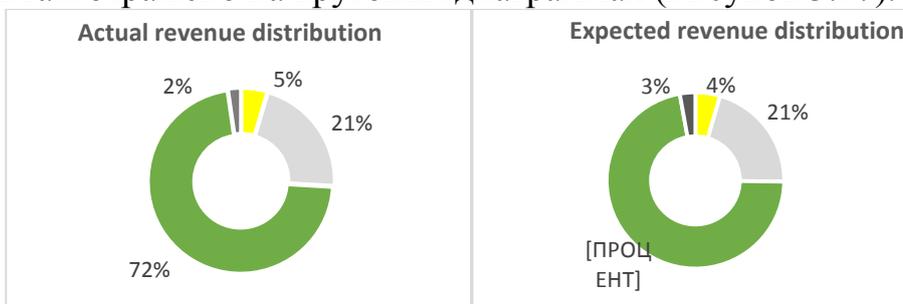
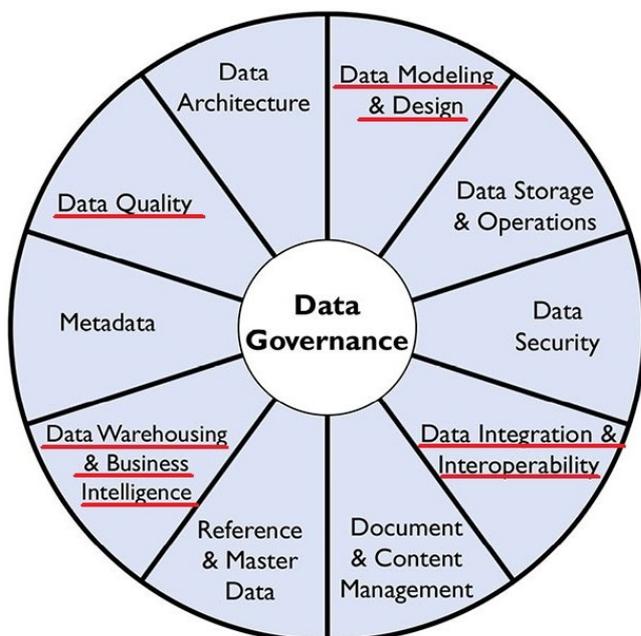


Рисунок 3.17 – Круговые диаграммы распределения клиентов по сегментам на основании скоринга: фактическая выручка (диаграмма слева), и ожидаемая выручка (диаграмма справа)

Примечание – Разработка автора

Важным элементом шестого этапа внедрения скоринга является оценка эффективности применяемых подходов к разработке. При реализации проекта по созданию скоринга применялись две методологии: DAMA-DMBOK, а также CRISP-DM.

На диаграмме ниже обозначены компоненты методологии DMBOK, которые были задействованы в ходе реализации проекта:



использовался на этапе подготовки данных, а также на этапе внедрения решения.

- Компонент Data Modeling & Design использовался в ходе изучения и моделирования данных;

- Компонент Data Quality использовался с целью проверки качества и валидности данных;

- Компонент Data Integration & Interoperability использовался в процессе внедрения новой процедуры скоринга в информационные системы компании;

- Компонент Warehousing & Business Intelligence

Выводы по главе 3

Рассмотренная поэтапная реализация разработки и внедрения новой процедуры скоринга демонстрирует актуальность применения технологии анализа больших данных в телекоммуникационной индустрии. Также практическим результатом является успешное применение компонентов методологии управления экосистемой данных DAMA-DMBOK.

На основании шестиэтапной методологии CRISP-DM было произведено структурирование процесса разработки аналитического решения. Данная методология может быть рекомендована для реализации проектов, связанных с анализом данных, в телекоммуникационных компаниях.

Внедрение новой процедуры скоринга повысило экономическую эффективность контентной бизнес-стратегии за счет оптимизации бизнес-процессов.

Главными результатами внедрения интеллектуального скоринга являются:

- возможность построения сегментации клиентов;
- возможность использования оценки ожидаемой выручки в целях развития медиа-сервиса;

- возможность применять результаты скоринга в целях оптимизации таргетированных маркетинговых кампаний;
- возможность эффективного управления программами лояльности в рамках медиа-сервиса;

На основании выделенных лучших практик управления ценностью данных, в частности шестиэтапного аналитического цикла CRISP-DM, был реализован аналитический проект по внедрению новой процедуры скоринга клиентов медиа-сервиса в компании velcom | A1.

ЗАКЛЮЧЕНИЕ

В работе проанализированы тенденции развития рынка телекоммуникаций. Важнейшей из них является внедрение телекоммуникационными компаниями новых конкурентных стратегий, основанных на данных. Также необходимо отметить, что формула лидерства, практикуемая телекоммуникационными компаниями на третьей фазе развития рынка и выражающаяся в контроле инфраструктуры, неизменно устаревает на четвертой фазе. Качество услуг и клиентский опыт, предоставляемый компаниями, являются важнейшими дифференциальными факторами на рынке. В качестве наиболее значимых современных трендов развития телекоммуникационной индустрии были выделены следующие: диверсификация проектного портфеля, поиск новых источников выручки, а также цифровая трансформация бизнеса.

В телекоммуникационной индустрии под влиянием интенсивного внедрения таких технологических подходов, как: аналитика больших данных, машинное обучение и искусственный интеллект, интернет вещей, а также роботизированная автоматизация процессов, происходит трансформация цепочек создания отраслевой ценности. Если ранее единственным источником ценности, создаваемой телеком-компаниями, были предоставляемые ими услуги, то постепенно произошло добавление нового источника ценности, основанного на подходах углубленной аналитики данных.

В работе раскрыты стратегии компенсации процессов коммодизации телекоммуникационных услуг компаниями рынка. Данная направленность обуславливается необходимостью внедрения инновационных бизнес-стратегий, способных коренным образом трансформировать восприятие конечного потребителя телеком-компаний. В связи с этим компании на рынке телекоммуникаций предпринимают усилия по превращению из провайдеров услуг в провайдеров конвергентных сервисов, интегрированных в единую экосистему. Примером таких экосистемных решений являются три рассмотренных бизнес-стратегии: контентная стратегия, предусматривающая создания медиа-платформы, предоставление data-продуктов, а также стратегия предоставления услуг финансового посредничества.

Важнейшим условием эффективного управления экосистемой данных телекоммуникационной компании является выбор абстрактной схемы формализации процессов управления данными. В результате сравнительного анализа методологий управления данными, в качестве наиболее применимой была определена методология DAMA-DMBOK. При рассмотрении подходов к извлечению ценности из данных, были проанализированы 3 методологии: SEMMA, KDD и CRISP-DM. CRISP-DM соответствует современным

практикам построения систем углубленной аналитики, в связи с чем, данный подход может быть рекомендован к внедрению телекоммуникационными компаниями в качестве методологической основы для извлечения ценности из данных. Помимо методологий управления данными и способами извлечения из них ценности, также были выделены организационные схемы, учитывающие цели построения высоко эффективной аналитической культуры. Учитывая ориентацию на данные в процессах принятия решений, наиболее эффективной является гибридная организационная модель.

В результате анализа подходов управления экосистемой данных был сформирован набор методологий, при использовании которых достигается наибольшая эффективность реализации потенциала данных.

На основании выделенных лучших практик управления ценностью данных, в частности шестиэтапного аналитического цикла CRISP-DM, был реализован аналитический проект по внедрению новой процедуры скоринга клиентов медиа-сервиса в компании velcom | A1.

На первом этапе аналитического цикла CRISP-DM производилась декомпозиция задачи. Были определены критерии успешности выполнения проекта, проведен предварительный анализ, произведена оценка рисков и был построен план реализации проекта, а также сформулированы требования к новой процедуре скоринга.

На втором этапе был произведен предварительный анализ данных.

На третьем этапе были подготовлены данные для последующего их использования в целях построения модели прогнозирования вероятности пролонгирования подписок.

На четвертом этапе производился отбор наиболее значимых предикторов, также была построена скоринговая модель и оценено ее качество.

На пятом этапе была произведена оценка общего качества модели, а также принято решение о ее внедрении.

На шестом этапе была произведена интеграция новой процедуры скоринга в информационные системы компании velcom | A1, а также была произведена автоматизация расчета скоринга.

Результаты исследования докладывались и обсуждались на научной конференции «76-я научная конференция студентов и аспирантов БГУ». Тезисы доклада приняты к публикации в материалах конференции.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. The Economic Impact of Telecommunications Networks and Broadband Internet: A Survey [Электронный ресурс]. – Режим доступа: <http://ftp.zew.de/pub/zew-docs/dp/dp16056.pdf>. – Дата доступа: 15.01.2019;
2. European Telecommunications Network Operators' Association [Электронный ресурс]. – Режим доступа: [https://etno.eu/datas/publications/annual-reports/ETNO%20Annual%20Economic%20Report%202017%20\(final%20version%20web\).pdf](https://etno.eu/datas/publications/annual-reports/ETNO%20Annual%20Economic%20Report%202017%20(final%20version%20web).pdf). – Дата доступа: 24.03.2019;
3. Stone M, The evolution of the telecommunications industry — What can we learn from it? / Journal of Direct, Data and Digital Marketing Practice, 2015.
4. Нервная система [Электронный ресурс]. – Режим доступа: https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D1%80%D0%B2%D0BD%D0%B0%D1%8F_%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0. – Дата доступа: 18.03.2019;
5. Аникин Борис Александрович, Аникин Олег Борисович Уровни развития мирового рынка телекоммуникаций // Вестник ГУУ. 2015. №2.
6. ICT Factors and Figures [Электронный ресурс]. – Режим доступа: <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf>. – Дата доступа: 12.03.2019;
7. The telecom price wars continue to rage in the global wireless industry [Электронный ресурс]. – Режим доступа: <https://www.strategyand.pwc.com/press/telco-price-wars-continue>. – Дата доступа: 12.03.2019;
8. IT – Telecom Business [Электронный ресурс]. – Режим доступа: <https://celikalper.wordpress.com/tag/gartners-hype-cycle/>. – Дата доступа: 12.03.2019;
9. Global telecommunications study: navigating the road to 2020 [Электронный ресурс]. – Режим доступа: [https://www.ey.com/Publication/vwLUAssets/Sektor_telekomunikacyjny/\\$FILE/ey_global_telecommunications_study_navigating_the_road_to_2020.pdf](https://www.ey.com/Publication/vwLUAssets/Sektor_telekomunikacyjny/$FILE/ey_global_telecommunications_study_navigating_the_road_to_2020.pdf). – Дата доступа: 17.03.2019;
10. Quality Productivity and Competitive Position [Электронный ресурс]. – Режим доступа: <https://www.amazon.com/Quality-Productivity-Competitive-Position-Edwards/dp/0911379002>. – Дата доступа: 12.03.2019;
11. Opportunities in Telecom Sector Arising from Big Data [Электронный ресурс]. – Режим доступа: <https://www2.deloitte.com/in/en/pages/technology-media-and->

telecommunications/articles/opportunities-in-telecom-sector-big-data.html. – Дата доступа: 12.03.2019;

12. Wikipedia: Большие данные [Электронный ресурс]. – Режим доступа:

https://ru.wikipedia.org/wiki/%D0%91%D0%BE%D0%BB%D1%8C%D1%88%D0%B8%D0%B5_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5.

– Дата доступа: 15.03.2019;

13. Rusabase: Что такое Big data [Электронный ресурс]. – Режим доступа: <https://rb.ru/howto/что-такое-big-data/>. – Дата доступа: 12.03.2019;

14. Gartner: Big Data [Электронный ресурс]. – Режим доступа: <https://www.gartner.com/it-glossary/big-data/>. – Дата доступа: 12.03.2019;

15. English Oxford Dictionary: Big Data [Электронный ресурс]. – Режим доступа: https://en.oxforddictionaries.com/definition/big_data. – Дата доступа: 12.03.2019;

16. Cambridge Dictionary: Big Data [Электронный ресурс]. – Режим доступа:

<https://dictionary.cambridge.org/ru/%D1%81%D0%BB%D0%BE%D0%B2%D0%B0%D1%80%D1%8C/%D0%B0%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9/big-data>. – Дата доступа: 12.03.2019;

17. Pattern Recognition and Machine Learning [Электронный ресурс]. – Режим доступа: <http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf>. – Дата доступа: 12.03.2019;

18. The Internet of Things: The next growth engine for the semiconductor industry [Электронный ресурс]. – Режим доступа: <https://www.pwc.com/gx/en/industries/technology/publications/internet-of-things.html>. – Дата доступа: 16.03.2019;

19. Beyond the dumb pipe: The IoT and the new role for network service providers [Электронный ресурс]. – Режим доступа: <https://www2.deloitte.com/insights/us/en/focus/internet-of-things/iot-in-telecom-industry.html>. – Дата доступа: 18.03.2019;

20. Robotic Process Automation — новый взгляд на старые технологии [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/company/jetinfosystems/blog/447398/>. – Дата доступа: 12.03.2019;

21. RPA FOR TELECOM [Электронный ресурс]. – Режим доступа: <https://www.automationanywhere.com/solutions/telecom>. – Дата доступа: 12.03.2019;

22. Arthur D. Little / Convergence of banking and telecoms, 2015;

23. Wikipedia: Виртуальный оператор сотовой связи [Электронный ресурс]. – Режим доступа: https://ru.wikipedia.org/wiki/%D0%92%D0%B8%D1%80%D1%82%D1%83%D0%B0%D0%BB%D1%8C%D0%BD%D1%8B%D0%B9_%D0%BE%D0%BF%D0%B5%D1%80%D0%B0%D1%82%D0%BE%D1%80_%D1%81%D0%BE%D1%82%D0%BE%D0%B2%D0%BE%D0%B9_%D1%81%D0%B2%D1%8F%D0%B7%D0%B8. – Дата доступа: 12.03.2019;
24. Data Management Body Of Knowledge [Электронный ресурс]. – Режим доступа: <https://dama.org/content/body-knowledge>. – Дата доступа: 12.05.2019;
25. Data Management Maturity (DMM) [Электронный ресурс]. – Режим доступа: <https://cmmiinstitute.com/data-management-maturity>. – Дата доступа: 12.03.2019;
26. McKinsey Quarterly: Data Culture [Электронный ресурс]. – Режим доступа: <https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20analytics/our%20insights/mckinsey%20quarterly%202018%20number%203%20overview%20and%20full%20issue/mckinsey-quarterly-2018-number-3.ashx>. – Дата доступа: 22.03.2019;
27. McKinsey: Why Data Culture Matters [Электронный ресурс]. – Режим доступа: <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/why-data-culture-matters>. – Дата доступа: 12.03.2019;
28. How to create a data culture [Электронный ресурс]. – Режим доступа: <https://www.cognizant.com/InsightsWhitepapers/how-to-create-a-data-culture-codex1408.pdf>. – Дата доступа: 12.03.2019;
29. CRISP-DM, still the top methodology for analytics, data mining, or data science projects [Электронный ресурс]. – Режим доступа: <https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>. – Дата доступа: 12.03.2019;
30. Data Mining Using SAS(R) Enterprise Miner(TM): A Case Study Approach, Third Edition [Электронный ресурс]. – Режим доступа: <http://support.sas.com/documentation/cdl/en/emcs/66392/HTML/default/viewer.htm#n0pejm83csbja4n1xueveo2uoujy.htm>. – Дата доступа: 12.03.2019;
31. SEMMA [Электронный ресурс]. – Режим доступа: <https://www.bigdataschool.ru/wiki/semma>. – Дата доступа: 12.03.2019;
32. KDD meets Big Data [Электронный ресурс]. – Режим доступа: <https://www.semanticscholar.org/paper/KDD-meets-Big-Data-Grady/bec7f9de0ec5bef2ccd1d4f518e0152fa80b9ed5>. – Дата доступа: 12.03.2019;
33. A survey of data mining and knowledge discovery process models and methodologies [Электронный ресурс]. – Режим доступа:

<https://www.cambridge.org/core/journals/knowledge-engineering-review/article/survey-of-data-mining-and-knowledge-discovery-process-models-and-methodologies/C2EC780B41545D44AB7F8F7BCBA8D982>. – Дата доступа: 12.03.2019;

34. Руководство по CRISP-DM для IBM SPSS Modeler [Электронный ресурс]. – Режим доступа: <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/18.0/ru/ModelerCRISPDm.pdf>. – Дата доступа: 12.03.2019;

35. Towards an Improved ASUM-DM Process Methodology for Cross-Disciplinary Multi-organization Big Data & Analytics Projects [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/326307750_Towards_an_Improved_ASUM-DM_Process_Methodology_for_Cross-Disciplinary_Multi-organization_Big_Data_Analytics_Projects_13th_International_Conference_KMO_2018_Zilina_Slovakia_August_6-10_2018_Proceeding. – Дата доступа: 16.03.2019;

36. Telcos: The untapped promise of big data [Электронный ресурс]. – Режим доступа: <https://www.mckinsey.com/industries/telecommunications/our-insights/telcos-the-untapped-promise-of-big-data>. – Дата доступа: 12.03.2019;

37. CDR analysis using Big Data technology [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/304292087_CDR_analysis_using_Big_Data_technology. – Дата доступа: 12.03.2019;

38. Mapping a future for the telecoms industry [Электронный ресурс]. – Режим доступа: <https://inform.tmforum.org/internet-of-everything/2017/05/mapping-future-telecoms-industry/>. – Дата доступа: 12.03.2019;

39. Top 5 ways the Telecom sector is leveraging Location Data [Электронный ресурс]. – Режим доступа: <https://www.dmtispatial.com/2018/07/25/top-5-ways-telecom-sector-leveraging-location-data/>. – Дата доступа: 17.03.2019;

40. Big Data and the Telecom Industry [Электронный ресурс]. – Режим доступа: <https://www.tatateleservices.com/downloads/WhitePapers/resources/Big-Data-and-the-Telecom-Industry.pdf>. – Дата доступа: 12.04.2019;

41. Guandong Xu, Web Mining and Social Networking: Techniques and Applications / Guandong Xu, Yanchun Zhang, Lin Li, Springer-Verlag Berlin, Heidelberg 2010;

42. Mishra, Sushruta & Mishra, Brojo & Tripathy, H & Mishra, Monalisa. (2017). Use of Social Network Analysis in Telecommunication Domain Use of Social Network Analysis in Telecommunication Domain;

43. Data Strategy [Электронный ресурс]. – Режим доступа: <https://www.ebooks.com/en-ag/95749240/data-strategy/bernard-marr/>. – Дата доступа: 12.03.2019;
44. Top 10 Data Science Use Cases in Telecom [Электронный ресурс]. – Режим доступа: <https://www.kdnuggets.com/2019/02/top-10-data-science-use-cases-telecom.html>. – Дата доступа: 03.03.2019;
45. Yildiz, Mumin & Varlı Albayrak, Songul. (2015). Customer churn prediction in telecommunication. 2015 23rd Signal Processing and Communications Applications Conference, SIU 2015 - Proceedings. 256-259.;
46. Analytical imperatives for telecom marketers in emerging markets [Электронный ресурс]. – Режим доступа: <https://www.strategyand.pwc.com/media/file/Hitting-the-target.pdf>. – Дата доступа: 12.03.2019;
47. Network Traffic Modeling for Load Prediction [Электронный ресурс]. – Режим доступа: https://www.mps.si/dokumenti/clanek_Svigelj_et_al.pdf. – Дата доступа: 12.03.2019;
48. Речевая аналитика колл-центров [Электронный ресурс]. – Режим доступа: <https://alpha-analytics.pro/>. – Дата доступа: 12.03.2019;
49. Чат-боты для индустрии телекоммуникаций [Электронный ресурс]. – Режим доступа: <https://www.chatinbox.io/chatbot-for-telecom/>. – Дата доступа: 12.03.2019;
50. Telecom customer segmentation based on cluster analysis [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/261389848_Telecom_customer_segmentation_based_on_cluster_analysis. – Дата доступа: 12.03.2019;
51. Modeling Customer Lifetime Value in the Telecom Industry [Электронный ресурс]. – Режим доступа: http://www.pm.lth.se/fileadmin/pm/Exjobb/Filer_fram_till_foerra_aaret/Exjobb_2013/Flordahl___Friberg/CLV_ERICSSON_Floldal_Friberg.pdf. – Дата доступа: 12.03.2019;
52. Andrae, Anders & ÖSTERMARK, ULF & Liu, Johan. (2000). Life Cycle Assessment of a telecommunications exchange. Journal of Electronics Manufacturing. 10.;
53. Top 10 Data Science Use cases in Telecom [Электронный ресурс]. – Режим доступа: <https://medium.com/activewizards-machine-learning-company/top-10-data-science-use-cases-in-telecom-e9c334d42616>. – Дата доступа: 12.03.2019;
54. An Investigation into Real-time Fraud Detection in the Telecommunications Industry [Электронный ресурс]. – Режим доступа: <https://paul.kinlan.me/telecom-fraud-detection/>. – Дата доступа: 12.03.2019;

55. Data-driven product development is more about process, culture, and people than technology [Электронный ресурс]. – Режим доступа: <https://snowplowanalytics.com/blog/2018/02/02/data-driven-product-development-is-more-about-process-culture-and-people-than-technology/>. – Дата доступа: 12.03.2019;
56. Титов А.И.. "Управление рисками ИТ-проектов на основе компонентной структуры разрабатываемого программного обеспечения" Интеллектуальные технологии на транспорте, no. 4, 2017, pp. 12-17;
57. Ciuca, Vasilica & Maer Matei, Monica. (2010). „Survival analysis for the unemployment duration” ;
58. Cox, David. (1972). Regression Models and Life Table. Journal of the Royal Statistical Society. Series B. 34;
59. NCSS Statistical System [Электронный ресурс]. – Режим доступа: http://ncss.wpengine.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Cox_Regression.pdf. – Дата доступа: 04.03.2019;
60. A Guide to Data Analysis and Critical Appraisal [Электронный ресурс]. – Режим доступа: <https://www.twirpx.com/file/1141677/>. – Дата доступа: 12.03.2019;
61. Understanding survival analysis: Kaplan-Meier estimate [Электронный ресурс]. – Режим доступа: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3059453/>. – Дата доступа: 10.02.2019;
62. Applied Survival Analysis: Regression Modeling of Time-to-Event Data [Электронный ресурс]. – Режим доступа: <https://www.wiley.com/en-gu/Applied+Survival+Analysis:+Regression+Modeling+of+Time+to+Event+Data,+2nd+Edition-p-9780471754992>. – Дата доступа: 12.03.2019;
63. IBM SPSS statistics 19 advanced statistical procedures companion [Электронный ресурс]. – Режим доступа: <https://lib.ugent.be/en/catalog/rug01:001802658>. – Дата доступа: 19.04.2019;
64. Modeling Customer Lifetime Value Using Survival Analysis [Электронный ресурс]. – Режим доступа: <https://support.sas.com/resources/papers/proceedings/proceedings/sugi28/120-28.pdf>. – Дата доступа: 24.02.2019.