

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
(МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)**

**МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ КОМИССИОННОГО ДОХОДА
КЛИЕНТСКОГО СЕГМЕНТА БАНКА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ
МАШИННОГО ОБУЧЕНИЯ**

2021

РЕФЕРАТ

Тема магистерской диссертации:

Моделирование и прогнозирование комиссионного дохода клиентского сегмента банка с использованием методов машинного обучения

Магистерская диссертация выполнена на 101 странице, содержит 20 таблиц, 48 рисунков, 62 использованных источника.

Актуальность темы диссертационной работы: необходимость создания научно-обоснованной методики планирования показателей эффективности коммерческого банка как полноправного участника финансового рынка.

Цель работы: теоретическое и методическое обоснование внедрения методов машинного обучения в систему планирования комиссионного дохода клиентского сегмента банка.

Поставлены следующие задачи:

- провести обзор практики применения методов машинного обучения в банковской сфере;
- изучить классификацию и особенности применения методов машинного обучения;
- построить модель бизнес-процесса AS-IS, выявить проблемы, провести их системный анализ;
- построить модель бизнес-процесса TO-BE;
- реализовать инструмент планирования комиссионного дохода клиентского сегмента банка на основе методов машинного обучения.

Объект исследования: методы и алгоритмы машинного обучения и их применение в задачах планирования ключевых показателей эффективности коммерческих банков.

Предмет исследования: разработка модели планирования комиссионного дохода клиентского сегмента коммерческого банка.

Научная новизна: разработана методика планирования отдельных показателей эффективности деятельности коммерческого банка.

Практическая значимость: предложенная в работе модель планирования принята Центральным аппаратом ПАО Сбербанк к пилотированию, что подтверждается Актом о внедрении научно-исследовательских разработок.

Экономическая эффективность: результаты внедрения модели подтверждены расчетами экономических показателей.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 ОБЗОР И СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ, МОДЕЛЕЙ И МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ	8
1.1 Применение методов машинного обучения в банковской сфере	8
1.2 Обзор существующих подходов к оценке комиссионного дохода клиентского сегмента банка.....	13
1.3 Классификация и сравнительный анализ инструментов машинного обучения.....	21
Результаты и выводы первой главы.....	27
2 ПОСТРОЕНИЕ РЕГРЕССИОННОЙ МОДЕЛИ ДЛЯ АНАЛИЗА КЛИЕНТСКОЙ БАЗЫ БАНКА.....	28
2.1 Построение полной бизнес-модели ПАО Сбербанк	28
2.2 Системный анализ проблемы	40
2.3 Разработка алгоритма построения регрессионной модели на основе деревьев решений	55
Результаты и выводы второй главы	65
3 РАЗРАБОТКА МОДЕЛИ ПЛАНИРОВАНИЯ КОМИССИОННОГО ДОХОДА КАНАЛА «СБЕРБАНК ПРЕМЬЕР».....	66
3.1 Постановка задачи и описание источника данных.....	66
3.2 Определение значимых критериев классификации для включения в деревья решений	68
3.3 Этапы обработки данных	75
3.4 Результаты работы модели и экономическое обоснование.....	79
Результаты и выводы третьей главы.....	93
ЗАКЛЮЧЕНИЕ.....	94
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	96

ВВЕДЕНИЕ

Одним из основных условий устойчивого развития рыночной экономики является стабильность финансового рынка, которая обеспечивается, в первую очередь, за счет системного функционирования и эффективного регулирования кредитных организаций – ключевых элементов банковской системы. Для этого в любом банке должна быть выстроена четкая система планирования и прогнозирования основных показателей его деятельности. Частью такой системы являются не только текущие, средне- и долгосрочные планы деятельности банка, но и стратегическое планирование, разработка направления развития банка, которая напрямую связана с планированием комплекса банковского маркетинга.

В периоды экономических кризисов задача поддержания надежного функционирования коммерческих банков стоит особенно остро, поскольку от этого зависит устойчивость финансового рынка в целом. В свою очередь, эффективное функционирование коммерческих банков как участников финансового рынка и его доминирующих секторов – денежного рынка, рынка ценных бумаг, рынка ссудного капитала – непосредственно зависит от обоснования текущего, средне- и долгосрочного планирования, а также от разработки стратегии развития.

В целом развитие розничного банковского бизнеса происходит в условиях острой конкуренции, которая ведет к появлению и активному распространению новых банковских продуктов и услуг, к повышению качества банковского обслуживания. В целях улучшения клиентского опыта проводится сегментация клиентской базы, под каждый сегмент создаются свои управленческие вертикали и разрабатываются стандарты работы. Кроме того, конкурентная борьба зачастую стимулирует коммерческие банки к существенной трансформации всего спектра своей деятельности, в том числе с использованием новейших достижений как в сфере финансово-кредитных операций, так и в области информационно-аналитической поддержки принятия управленческих решений.

Сегодня данные являются одним из самых дорогих банковских активов. Банки имеют доступ к огромному количеству данных, занимаются их обработкой и хранением. Однако для того, чтобы получить конкурентное преимущество, необходимо преобразовывать накопленные массивы данных в полезные знания, которые в обобщенном и информативном виде должны быть предоставлены руководству. В этой связи становится понятным стремление финансовых организаций к использованию экономико-математических методов и инструментов бизнес-аналитики.

С точки зрения принятия управленческих решений в настоящее время наиболее востребовано такое направление информационных технологий, как интеллектуальный анализ данных, или Data Mining (DM). С английского языка этот термин можно перевести как «разработка, раскопка данных». Это междисциплинарное направление, которое возникло на стыке математической статистики, машинного обучения и искусственного интеллекта, активно применяется для решения задач прогнозирования, классификации и кластеризации.

DM не предлагает готовых алгоритмов решения для той или иной задачи анализа. Он представляет собой совокупность приемов и методов аналитической обработки данных, которые позволяют извлекать из них максимум полезных знаний. Ядро аналитических технологий DM составляют методы машинного обучения. С их помощью эксперт или аналитик получает возможность автоматически находить закономерности в данных и определять зависимости между различными признаками, после чего он может делать выводы о состоянии и развитии изучаемых явлений и процессов и выработать рекомендации по дальнейшей работе с ними.

Проекты по разработке и внедрению DM-технологий в практическую деятельность предприятий и организаций в целях повышения качества принятия управленческих решений отличаются высокой стоимостью, трудоемкостью и требуют определенного уровня профессиональных компетенций. В числе основных проблем выделяются отсутствие формальной постановки задачи,

плохо проработанная стратегия поиска знаний, а также низкое качество исходных данных. Поэтому разработка новых подходов и методов по реализации DM-проектов в социально-экономических системах является важной задачей, требующей решения.

Актуальность диссертационного исследования заключается в том, что в настоящее время существует острая необходимость в разработке научно-обоснованной методики планирования ключевых показателей деятельности коммерческого банка как самостоятельного субъекта финансового рынка.

Основной целью диссертации является теоретическое и методическое обоснование внедрения методов машинного обучения в систему планирования комиссионного дохода клиентского сегмента банка.

В соответствии с основной целью диссертационной работы определены следующие задачи:

- провести обзор практики применения методов машинного обучения в банковской сфере;
- сформулировать теоретические и методологические особенности инструментов, применяемых в планировании и прогнозировании финансовых результатов деятельности коммерческих банков;
- рассмотреть классификацию и изучить особенности применения методов машинного обучения;
- построить модель существующего бизнес-процесса AS-IS, который требует совершенствования, выявить проблемы, провести их системный анализ;
- на основании проведенного анализа провести совершенствование бизнес-процесса, построить модель TO-BE;
- разработать алгоритм реализации практического инструмента планирования комиссионного дохода клиентского сегмента банка;
- произвести оценку экономической эффективности проекта по внедрению предлагаемого инструмента.

Объектом исследования выступают методы и алгоритмы машинного обучения и их применение в задачах планирования ключевых показателей эффективности коммерческих банков.

Предметом исследования является разработка модели планирования комиссионного дохода клиентского сегмента коммерческого банка.

Источником данных для исследования служат ретроспективные данные по клиентской базе высокодоходного клиентского сегмента ПАО Сбербанк.

Теоретическую и методологическую базу диссертационной работы составляют современные труды в области менеджмента, планирования и регулирования деятельности коммерческих организаций, управления кредитной организацией, интеллектуального анализа данных и машинного обучения.

В диссертации использованы общенаучные методы исследования, такие как анализ, синтез, дедукция и индукция, а также методы прогностического, логико-аналитического и математического моделирования, математической статистики, прикладной статистики, искусственного интеллекта, машинного обучения, обработки больших данных, кластерного анализа.

В работе использованы результаты исследований и практической деятельности автора, основные теоретические положения и выводы современной экономической науки, а также обобщенные результаты исследований в области интеллектуального анализа данных и машинного обучения.

Научная новизна диссертационной работы состоит в том, что в ходе исследования разработана методика планирования отдельных показателей эффективности коммерческого банка, которая может применяться для совершенствования бизнес-процессов планирования в различных кредитных организациях.

Практическая значимость работы заключается в разработке модели планирования комиссионного дохода клиентского сегмента банка с использованием методов машинного обучения. Предложенная модель принята к пилотированию в ПАО Сбербанк, что подтверждается актом о внедрении

научно-исследовательских разработок. По итогам пилотирования будет принято решение о возможности централизованного внедрения модели в работу банка.

Автором в рамках улучшения качества системы планирования было предложено использование инструментов с использованием методов машинного обучения в процессе бизнес-планирования финансовых результатов банка и доказана экономическая эффективность внедрения таких инструментов в настоящую систему на примере прогнозирования комиссионного дохода канала «Сбербанк Премьер» ПАО Сбербанк.

Также отдельные положения и выводы диссертационного исследования реализованы в практике деятельности ПАО Сбербанк при совершенствовании системы бизнес-планирования финансовых результатов.

Диссертационная работа имеет структуру, включающую в себя введение, три главы, разделенные на 10 параграфов, заключение, список использованной литературы. Работа содержит 20 таблиц и 48 рисунков.

Таблицы, рисунки, диаграммы и схемы, представленные в тексте диссертационной работы, выполнены автором с помощью программных продуктов Microsoft Excel, Microsoft Visio, Microsoft Project, ArisExpress.

Первая глава диссертационной работы отражает теоретические основы методов машинного обучения и описание опыта их практического применения в деятельности коммерческих банков. Вторая глава посвящена анализу деятельности ПАО Сбербанк, исследованию существующей системы планирования комиссионного дохода «Сбербанк Премьер», поиску недостатков в существующей системе и разработке путей совершенствования обнаруженных проблем. Третья глава описывает этапы практической реализации предлагаемой модели планирования комиссионного дохода «Сбербанк Премьер», а также экономическое обоснование ее внедрения.

1 ОБЗОР И СРАВНИТЕЛЬНЫЙ АНАЛИЗ ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ, МОДЕЛЕЙ И МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

1.1 Применение методов машинного обучения в банковской сфере

Термин «машинное обучение» в 1959 г. ввел в научный обиход А. Сэмюэл, американский исследователь в области компьютерных игр и искусственного интеллекта [55]. Как научная дисциплина машинное обучениеросло из стремления к искусственному интеллекту (ИИ). Первые исследователи были заинтересованы в том, чтобы машины учились на данных и пробовали подойти к проблеме с помощью первых прототипов «нейронных сетей»: это были, главным образом, перцептроны [37] и другие модели, которые позже были признаны в обобщенных линейных моделях статистики.

В 70-80-е гг. XX в. в рамках машинного обучения были предложены деревья решений [54], ассоциативные правила [51] и самоорганизующиеся карты признаков [53]. Значительный вклад в области моделирования социальных и экономических систем с целью анализа их функционирования и синтеза управленческих решений внесли отечественные ученые [8]. В современных исследованиях российских авторов рассматриваются способы применения инструментов машинного обучения в различных социально-экономических системах для решения прикладных задач.

Так, предметом изучения в работе [33] являются методы и алгоритмы DM, методология и проблемы их применения в задачах моделирования объектов и процессов в экономической, социальной и бизнес среде. В частности, разрабатываются критерии и рекомендации для выбора DM-средств с точки зрения внедрения на уровне специалистов, непосредственно интегрированных в процессы в социальных и экономических системах.

В рамках [34] предлагаются сценарии построения интеллектуальных моделей, основанных на машинном обучении, с использованием декомпозиции процесса моделирования на этапы.

Данные подходы применяются для построения комплексной интеллектуальной модели, использующей деревья решений для анализа клиентской базы кредитной организации с целью совершенствования маркетинговой стратегии на основе исследования зависимости свойств клиента и его отклика на коммерческие предложения [10].

В работе [44] на основе технологий Big Data разрабатывается информационная модель социального профиля человека, которая учитывает различные форматы представления, источники и динамику изменения открытых социальных данных.

Эти разработки получили развитие в другой работе за тем же авторством [46], в которой на основе методов анализа гетерогенных данных, описанных в [45], предложена математическая модель цифровой социальной среды, отличающаяся многоуровневым описанием ее структуры, включающим иерархию социальных групп, отдельных личностей и социальных явлений. Модель позволяет учитывать явные и неявные связи между элементами социальной среды с возможностью последующей алгоритмизации аналитических функций для поддержки лиц, принимающих решения.

В целях обеспечения соответствия данных элементам предложенной информационной модели социальной среды, в рамках [43] разработаны алгоритмы фильтрации и ранжирования данных социального профиля, а также выявления типажей с различными настройками в разрезе территорий или социальных групп для мониторинга социальной напряженности и оценки мнений и реакций людей на определенные события. Отличием данных алгоритмов является комплексное использование совокупности тезаурусов, методов лингвистического анализа текстовой информации, а также средств извлечения семантики из мультимедиа данных, характеризующих объекты цифровой социальной среды.

Работы [31], [57] и [32] посвящены исследованию экономико-математического, алгоритмического и программного обеспечения анализа финансовых рынков.

В частности, в [31] разработан метод генерации выходной выборки для обучения нейронной сети с учетом повышения волатильности, использующийся в задаче классификации рыночных ситуаций. Данный метод отличается возможностью задания ожидаемой прибыли и позволяет исключить влияние квалификации эксперта на эффективность работы интеллектуальной системы поддержки принятия решений и учесть риск повышения волатильности, связанный с влиянием макроэкономических показателей и новостей.

Методика поддержки принятия решений трейдера по направлению сделки с финансовым инструментом, разработанная в [57], заключается в нейросетевой классификации рыночных ситуаций с применением методов технического, фундаментального анализа и эконометрического моделирования. Предложенная методика позволяет проводить анализ многокритериальной информации о текущем состоянии рынка конкретного финансового инструмента, учитывая основные подходы к анализу финансовых рынков: прогнозирования временных рядов, технического анализа, фундаментального анализа (в части анализа влияния на повышение волатильности выхода экономических новостей).

Результатом этих разработок становится реализация интеллектуальной системы поддержки принятия решений: в [32] описывается функциональное и детальное проектирование системы, программное, информационное и математическое обеспечение системы, а также результаты ее экспериментального использования. Предложенная система отличается возможностью задания параметров анализа в виде ожидаемой прибыли от сделки с конкретным финансовым инструментом на выбранном временном интервале, наличием модуля фундаментального анализа, позволяющего выявлять для выбранного финансового инструмента значимые новости и события из экономического календаря, а также кэша данных, позволяющего обращаться к предыдущим результатам анализа для выбранного финансового инструмента, сокращая таким образом затраты при повторной подготовке данных системы для нейросетевого анализа. Все результаты обучения нейронной сети для различных финансовых инструментов и временных интервалов хранятся в базе данных, что

также снижает время обучения нейронной сети при многократном использовании системы для повторяющихся параметров.

В работах, связанных непосредственно с банковским сектором, разработана концептуальная модель процесса оценки платежеспособности заемщиков [22], в [23] разработана ее структура, особенностью которой является применение комбинированной базы знаний на основе модифицированного гипертекста, который декомпозирован на два структурных компонента, отражающих четкую и нечеткую информацию с применением дерева принятия решений [21]. Данная база знаний используется в качестве основной подсистемы информационной поддержки кредитных экспертов.

В исследовании [35] разработана базовая модель оценки управляющих параметров развития розничного бизнеса коммерческого банка. Развивая это направление, тот же автор в [15] предлагает модификации базовой модели, позволяющие определить перспективную структуру клиентской базы коммерческого банка с учетом эффективности различных категорий физических лиц. В результате сформирована комплексная процедура принятия управленческих решений, в рамках которой предложен подход для анализа конкурентных сил в сфере розничного банковского сектора с помощью методов кластерного анализа и метода самоорганизующихся карт-признаков Кохонена [14], определена методика сегментирования и оценки доходности операций по группам частных клиентов банка [13], а также выработаны рекомендации по выбору инструментального средства и способу его внедрения в условиях конкретного коммерческого банка [12].

Обзор практики применения инструментов интеллектуального анализа данных в различных социально-экономических системах показывает, что банковская система является одной из самых перспективных сфер для внедрения такого рода инструментов.

Машинное обучение оказывается эффективным на различных уровнях принятия решений. На уровне организации в целом его применяют для анализа конкурентной среды и комплексной оценки основных параметров развития

бизнеса. На уровне направлений бизнеса (инвестиционный, корпоративный, розничный бизнес) его используют в разработке интеллектуальных систем поддержки принятий решений. На уровне отдельного бизнес-блока, например, в розничном бизнесе, с его помощью решают узкоспециализированные практические задачи: сегментирование клиентов, оценка доходности операций по группам клиентов, планирование отдельных показателей бизнеса.

Эффективность применения методов машинного обучения позволяет автору сделать вывод о возможности их использования для построения модели планирования комиссионного дохода клиентского сегмента банка.

Важным вопросом для машинного обучения является качество данных. Данные должны подбираться индивидуально под каждую задачу и представлять собой набор значимых и информативных признаков, на основе которых будет моделироваться целевая переменная. Например, для построения модели планирования комиссионного дохода клиентского сегмента коммерческого банка такими признаками могут быть возраст и пол клиента, территория, на которой он обслуживается, остатки на его счетах и объем средств, вложенных в тот или иной банковский продукт и др.

Правильно подобранный инструмент машинного обучения в совокупности с адекватными данными позволяет эффективно решить задачу и принести положительный экономический эффект.

Использование инструментов DM в различных социально-экономических системах, и в частности, в банках, – это уже глобальный тренд. Согласно результатам международного исследования Microsoft, 94% руководителей считают, что технологии искусственного интеллекта важны для решения стратегических задач их организаций. При этом 27% опрошенных уже внедрили соответствующие технологии в ключевые бизнес-процессы, еще 46% ведут пилотные проекты [9, с. 42]. Многие организации по всему миру осваивают интеллектуальный анализ данных, поскольку этот подход позволяет эффективно распределять ресурсы, повысить производительность и принимать качественные управленческие решения.

1.2 Обзор существующих подходов к оценке комиссионного дохода клиентского сегмента банка

Планирование деятельности банка – сложный аналитический процесс, требующий от экономиста большого опыта и знаний для учета всех возможных факторов, влияющих на деятельность кредитной организации. Необходимость точного определения будущей социально-экономической и демографической ситуации в регионе, а также динамики показателей деятельности банка обуславливает широкое использование разных методов планирования и прогнозирования в зависимости от экономической сущности объекта исследования [27, с. 384].

Понятия планирования и прогнозирования тесно связаны между собой. Они выступают в качестве взаимодополняющих процессов при описании перспектив развития отдельных объектов, явлений или целых систем. Однако между ними существует одно главное различие: прогноз используется для тех показателей, которые находятся вне сферы влияния банка (уровень доходов населения, инфляция, темпы экономического роста и другие показатели макроэкономики), тогда как план рассчитывается для тех показателей, которыми банк может управлять (комиссионный доход, прирост клиентской базы, количество выданных кредитов и т.д.). Тем не менее, исходя из специфики банковской деятельности, внутри кредитной организации существует множество факторов, которые не зависят от нее в полной мере, например, денежно-кредитная политика и валютное регулирование, проводимые Центральным Банком, а также выполнение им надзорных функций. Эти факторы создают условия для широкого использования методов прогнозирования.

Универсального метода прогнозирования не существует из-за разнообразия моделируемых ситуаций. Экономисты используют свыше 100 методов прогнозирования. Один из вариантов их классификации, основанный на индуктивном и дедуктивном подходах, предполагает, что вся совокупность методов прогнозирования представлена двумя группами в зависимости от

степени их однородности: простые и комплексные. Группа простых методов объединяет однородные по содержанию и инструментарию методы прогнозирования. Комплексные методы – совокупности методов, реализуемые специальными прогностическими системами [19, с. 41].

Кроме того, все методы прогнозирования делятся на три класса: фактографические, экспертные (интуитивные) и комбинированные. В основу их выделения положен характер информации, на базе которой составляется прогноз:

1. *Фактографические методы* базируются на фактическом информационном материале о прошлом и настоящем развитии объекта прогнозирования. Чаще всего они применяются при поисковом прогнозировании для эволюционных процессов.
2. *Экспертные (интуитивные) методы* основаны на использовании знаний специалистов/экспертов об объекте прогнозирования и обобщении их мнений о развитии объекта. Данные методы, как правило применяются для прогнозирования волатильных процессов с большим количеством выбросов. Такие прогнозы, как правило, основаны на экспертных знаниях конкретного сотрудника или команды, а также обладают недостаточной степенью точности, поэтому не находят широкого применения в банковской деятельности. В основном их использование сводится к краткосрочному планированию второстепенных драйверов.
3. *Комбинированные методы* включают в себя методы со смешанной информационной основой, в которых в качестве первичной информации наряду с экспертной используется и фактографическая [24, с. 222].

Наиболее распространены в банковской деятельности следующие методы прогнозирования:

I. Статистические методы.

1. *Экстраполяция по скользящей средней*. Применяется для целей краткосрочного прогнозирования (до одного года). Необходимость в этом методе возникает в том случае, когда имеющиеся данные динамического ряда

не позволяют обнаруживать какую-либо тенденцию развития (тренд) того или иного процесса из-за случайных и периодических колебаний исходных данных.

Этот метод состоит в замене фактических уровней динамического ряда расчетными, имеющими значительно меньшую колеблемость, чем исходные данные. Средняя рассчитывается по группам данных за определенный интервал времени, причем каждая последующая группа образуется со сдвигом на один период.

Метод скользящей средней предполагает, что при вычислении средние как бы скользят от одного периода к другому; с каждым новым шагом средняя обновляется за счет новой информации о фактически реализуемом процессе. Таким образом, при прогнозировании исходят из предположения о том, что следующий во времени показатель по своей величине будет равен средней, рассчитанной за последний интервал времени.

Достоинством данного метода является простота использования, недостатком – невысокая степень вероятности реализации запланированного показателя (а для экономических процессов такая вероятность должна быть равна не менее 95%).

2. *Экспоненциальная средняя.* При расчете прогноза с помощью скользящей средней, чем раньше по времени наблюдение, тем меньше оно оказывает влияния на величину скользящей средней, т.е. влияние прошлых значений должно затухать по мере удаления от момента, для которого определяется средняя. В основе расчета экспоненциальной средней лежит прогноз линейной комбинации прошлых и текущих наблюдений. Этот прием позволяет сгладить динамический ряд, уменьшая влияние более старых периодов с помощью коэффициентов.

Данный метод обладает всеми преимуществами метода скользящей средней, но при этом использование сглаживающего коэффициента позволяет добиться большей точности прогноза. Однако функция экспоненциальной средней является линейной, что не позволяет использовать ее для явлений с

более сложными функциональными зависимостями. Несмотря на это, данный метод широко применяется для краткосрочного прогнозирования.

3. *Авторегрессионные зависимости.* В основу этого метода положена предпосылка о том, что экономические процессы имеют определенную специфику: они отличаются взаимозависимостью и инерционностью. Последняя означает, что значение практически любого экономического показателя в момент времени t зависит определенным образом от состояния этого показателя в предыдущих периодах: значения прогнозируемого показателя в прошлых периодах должны рассматриваться как факторные признаки.

Модели прогнозирования, основанные на авторегрессионном методе, являются на сегодняшний день популярным инструментом и используются для решения задач прогнозирования временных рядов различных предметных областей. Например, они часто применяются на финансовых рынках, где с их помощью моделируется волатильность.

К числу достоинств данного метода можно отнести простоту и прозрачность моделирования. Недостатками являются большое число параметров модели, идентификация которых неоднозначна и ресурсоемка, низкая адаптивность моделей, а также их линейность, вследствие чего их невозможно применять для моделирования нелинейных процессов, которые на практике встречаются гораздо чаще линейных зависимостей.

4. *Многофакторный регрессионный анализ.* Данный метод учитывает существующие связи между целевым показателем и другими факторами. Для их определения проводится качественный анализ исходных данных, в ходе которого выделяется несколько факторов, влияющих на изменение прогнозируемого показателя. После этого строится регрессионная зависимость определенного типа: линейная, квадратичная, гиперболическая, экспоненциальная и т.д.

Сильной стороной данного метода является высокая степень точности прогнозов, однако он предъявляет высокие требования к отбору наиболее

значимых факторов, влияющих на целевую переменную, поскольку они одновременно должны быть тесно связаны с моделируемым фактором и не являться взаимозависимыми, иначе это приведет к искажению модели. К недостаткам также можно отнести и сложность определения типа регрессионной зависимости, поскольку показатели банковской деятельности редко характеризуются линейными связями.

- II. Математический метод прогнозирования. Он необходим для того, чтобы формализовать (т.е. записать в терминах математической модели) задачи оптимального управления и построения прогноза.

Преимуществом использования данного метода является высокая точность планирования и прогнозирования. Однако этот метод не распространен вследствие сложности подбора показателей для векторов результирующих, статусных и управляемых переменных, а также характеристик внешней среды так, чтобы они были взаимосвязаны и сопоставимы. К тому же не все банковские процессы можно описать математическими формулами и при этом на определенных отрезках времени или при некоторых значениях переменных не потерять экономический смысл явления.

- III. План-прогноз на основе модели экономической динамики банка. Модель строится на основе теории и методов автоматического регулирования и учитывает влияние основных агрегированных экономических показателей банка, а также налоговых платежей и амортизации основных фондов на траекторию роста капитала банка. При моделировании делают принципиальное допущение о непрерывности финансовых потоков поступлений и платежей, которое оправдано в случае выбора достаточно длительного периода усреднения, например, месяца [50, с. 432].

Динамическая модель используется в расчетах прогнозов и планов для агрегированных показателей банковской деятельности. План-матрицы, разработанные на ее основе, позволяют рассчитывать плановые значения показателей на среднесрочный период в разрезе месяцев. Применяется следующая методика расчета: часть показателей модели задается в качестве

исходных параметров, а остальные рассчитываются с помощью уравнений модели. К примеру, если задать в качестве исходных данных объем привлеченных средств, рассчитанный методом экстраполяции или любым другим, то с помощью модели можно получить прогнозные значения для ежемесячной прибыли или размера собственного капитала. Такой расчет позволит наглядно продемонстрировать будущую динамику капитала при заданной динамике пассивов.

Преимуществом данного подхода является наглядность и высокая скорость моделирования возможных будущих ситуаций. Данные модели могут служить удобным прикладным инструментом как для рядового аналитика, так и для лица, принимающего решения. Основным недостатком заключается в сложности применения, поскольку планирование и прогнозирование подразумевает использование большого набора исходных параметров для увеличения точности модели.

Технологическое прогнозирование подразделяется на изыскательское и нормативное [18, с. 232].

Основу *изыскательского прогнозирования* составляет поиск трендов, тенденций развития явлений и процессов, исходя из накопленного массива данных, на котором строится прогноз. Изыскательское прогнозирование представляет собой прогрессивное движение в пространстве технологий от более примитивных методов и инструментов к высокотехнологичным.

Нормативное прогнозирование ориентируется на потребности и цели кредитной организации, к которым она стремится, и на ее миссию, которой она следует. Нормативное прогнозирование представляет собой регрессивное движение в пространстве технологий от высокотехнологичных инструментов к более упрощенным, низкоуровневым технологиям.

Технологическое прогнозирование применяется для решения задач расчета прогнозов в области социально-экономической и политической деятельности, а также коммерческой активности.

Поиск наиболее эффективного сочетания методов изыскательского и нормативного прогнозирования относится к числу основных проблем при разработке методик прогнозирования.

Для изыскательского прогнозирования характерно использование таких методов, как экстраполяция, моделирование, метод исторической аналогии, написание сценариев и т.д., базирующихся на анализе точных эмпирических данных [4, с. 149].

При использовании методов изыскательского прогнозирования предпочтение отдается количественной информации, хотя использование неколичественной информации в изыскательском прогнозировании также возможно.

Примером тому является использование интуитивных методов, метода сценариев или метода экспертных кривых, позволяющих определить наметившиеся тенденции изменения ситуации, базируясь не только на эмпирических данных, но и на опыте высококвалифицированных специалистов-экспертов.

К числу основных методов, используемых при нормативном прогнозировании, следует отнести методы ПАТТЕРН, Делфи, прогнозного графа Глушкова, Поспелова и др. [38, с. 52-58].

Среди других видов прогнозирования иногда выделяют прогнозирование с использованием обратной связи, интуитивные методы, обходные и т.д. Но основные идеи, используемые при разработке прогнозов, достаточно полно представлены именно в изыскательском и нормативном прогнозировании.

Обзор и сравнительный анализ существующих методов прогнозирования, их преимуществ и недостатков позволяет автору сделать вывод о том, что некоторые из них, например, статистические, могут эффективно применяться для планирования отдельных драйверов, использующихся в моделировании ключевых показателей эффективности. Однако для построения комплексной модели планирования конкретного показателя, включающей в себя множество параметров, целесообразно рассмотреть более мощные инструменты.

Активно развивающаяся область искусственного интеллекта – машинное обучение – включает в себя модели, методы и алгоритмы, ориентированные на автоматическое накопление и формирование знаний на основе анализа и обобщения данных.

Различают два вида обучения:

- индуктивное обучение, или обучение по прецедентам, базируется на выявлении эмпирических закономерностей в данных;
- дедуктивное обучение предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний.

Большинство методов обучения по прецедентам разрабатывалось как альтернатива классическим статистическим подходам [48, с. 40].

Основную задачу машинного обучения можно сформулировать следующим образом. Исходный массив данных содержит в себе множество объектов или ситуаций и множество возможных ответов или откликов на ситуацию. Объекты и ответы находятся в некоторой зависимости, природа которой неизвестна. Дана также результирующая совокупность соответствий «объект–ответ», которая называется обучающей выборкой. Задача заключается в том, чтобы, используя эти условия, определить характер зависимости. Иными словами, необходимо построить алгоритм, который будет способен определить достаточно точный ответ для любого вхождения объекта. При этом важно, чтобы обучаемая система была способна к обобщению, то есть ее аналитические возможности не должны ограничиваться обучающей выборкой, а давать адекватные ответы на любой входящий объект. Оценка точности ответов дается с помощью расчета определенных показателей качества модели.

Таким образом, описанная задача сводится к классическим задачам аппроксимации функций, объектами в которых являются действительные числа или векторы. На практике в исходных массивах данных зачастую содержатся неполные или неточные данные, информация может быть как числовой, так и категориальной. Все эти особенности приводят к появлению большого количества разнообразных методов машинного обучения.

1.3 Классификация и сравнительный анализ инструментов машинного обучения

Обобщая результаты работ современных исследователей в области машинного обучения [7, с. 30], [30, с. 7–8], [48, с. 25–26], можно сделать вывод, что все задачи, решаемые с его помощью, можно отнести к одной из следующих категорий:

1. *Задача регрессии* – прогноз на основе выборки объектов с различными признаками. На выходе получается вещественное число, к примеру, цена квартиры, стоимость ценной бумаги по прошествии полугода, ожидаемый комиссионный доход.
2. *Задача классификации* – определение категории объекта на основе набора признаков. Имеет конечное количество ответов (как правило, в формате «да» или «нет»). К такой задаче относится, например, решение о выдаче кредита.
3. *Задача кластеризации* – распределение данных на группы: разделение всех клиентов мобильного оператора по уровню платёжеспособности или отнесение клиента к тому или иному сегменту.
4. *Задача уменьшения размерности* – сведение большого числа признаков к меньшему (обычно 2–3) для удобства их последующей визуализации (например, сжатие данных).
5. *Задача выявления аномалий* – отделение аномалий от стандартных случаев. На первый взгляд она совпадает с задачей классификации, но есть одно существенное отличие: аномалии – явление редкое, и примеров, на которых можно обучить модель на выявление таких объектов, крайне мало, поэтому методы классификации в данном случае не применимы. На практике такой задачей является, например, выявление мошеннических действий с банковскими картами.

В публикациях по интеллектуальному анализу данных [11, с. 18], [29, с. 30], [52, с. 29] выделяются следующие способы машинного обучения:

1. *Обучение с учителем (Supervised learning)*. Этот способ оптимален, если имеется достаточно большая выборка данных, на которых можно обучать алгоритм и настраивать его параметры. Качество обучения можно оценить на контрольных данных, для которых строится прогноз. Чаще всего обучение с учителем применяется для задач классификации и прогнозирования. Данный способ можно использовать при определении финансового риска частных лиц и организаций на основе имеющихся сведений о прошлой финансовой активности. Другим примером является прогнозирование покупательского поведения на основе исторических данных о спросе на тот или иной товар.
2. *Обучение без учителя (Unsupervised learning)*. При обучении без учителя машина исследует набор данных и выявляет скрытые закономерности корреляции между различными переменными. Этот способ можно использовать для группирования данных в кластеры на основании одних только их статистических свойств. Хорошее примером применения обучения без учителя является алгоритм кластеризации, используемый для вероятностного соединения записей, когда определяются связи между элементами данных, и на основании этих отношений выявляются связи между людьми и организациями в физическом или виртуальном мире. Такой вариант особенно полезен компаниям, которым нужно, например, объединить данные из разнородных источников или по различным структурным подразделениям, чтобы построить общую картину клиентуры. Обучение без учителя также можно использовать для анализа тональности высказываний, чтобы определять эмоциональное состояние людей на основе их постов в социальных сетях, сообщений электронной почты и других записей. Кроме того, с помощью обучения без учителя в настоящее время все чаще оценивают уровень удовлетворенности клиентов в различных компаниях, в том числе в банках.
3. *Обучение с подкреплением (Reinforcement learning)*. При обучении с подкреплением машине позволяют взаимодействовать с окружением (например, сбрасывать бракованную продукцию с конвейера в корзину) и

«вознаграждают», когда она правильно выполняет задание. Автоматизировав подсчет вознаграждений, можно дать возможность машине обучаться самостоятельно. Одним из применений обучения с подкреплением может быть сортировка товаров в розничных магазинах. Некоторые продавцы экспериментируют с роботизированными системами сортировки предметов одежды, обуви и аксессуаров. Роботы, используя обучение с подкреплением, определяют, насколько сильно нужно сдвинуть предмет при хватании и какой хват будет наилучшим.

Некоторые специалисты по интеллектуальному анализу данных дополнительно выделяют и другие способы машинного обучения [62]:

1. *Обучение с частичным привлечением учителя (Semi-supervised learning)*. Это гибрид обучения с учителем и без. Разметив небольшую часть данных, учитель дает машине понять, каким образом кластеризовать оставшиеся данные. Способ можно применять для распознавания мошенничеств с попытками выдать себя за другого. Мошенничества можно классифицировать как аномалию на фоне обычной активности. Методы машинного обучения с частичным привлечением учителя позволяют создавать модели, распознающие такие аномалии. Соответствующие системы нередко применяются для выявления попыток мошенничеств при онлайн-сделках. Обучение с частичным привлечением учителя также можно применять в случаях, когда есть готовые наборы частично размеченных данных, что характерно для крупных предприятий. Например, в Amazon улучшили способности цифрового ассистента Alexa понимать естественный язык, обучая алгоритмы искусственного интеллекта на сочетаниях размеченных и неразмеченных данных. Благодаря этому удалось повысить точность ответов Alexa.
2. *Глубинное обучение (Deep learning)*. Глубинное обучение может проходить как без учителя, так и с подкреплением. При глубинном обучении частично имитируются принципы обучения людей, т.е. используются нейронные сети для все более подробного уточнения характеристик набора данных.

Глубинные нейронные сети применяются, в частности, для ускорения скрининга больших объемов данных при поиске лекарственных средств. Такие нейросети способны обрабатывать множество изображений за короткое время и извлечь больше признаков, которые модель в конечном счете запоминает. Этот способ также широко применяется для борьбы с мошенничествами, т.к. улучшает точность распознавания благодаря автоматизации. Глубинное обучение может использоваться в автомобильной отрасли при выполнении ремонта и технического обслуживания.

Дальнейший анализ публикаций [62], [17, с. 108], [6, с. 134, 200, 235] позволяет определить ряд основных алгоритмов, использующихся в моделях машинного обучения:

1. *Дерево решений*. Алгоритм дерева решений классифицирует объекты, отвечая на «вопросы» об их атрибутах, расположенные в узловых точках. В зависимости от ответа выбирается одна из ветвей, пока не будет достигнут «лист» – окончательный ответ. Среди применений дерева решений – платформы управления знаниями для клиентского обслуживания, прогнозного ценообразования и планирования выпуска продукции. В страховой компании дерево решений поможет выяснить, какие виды страховых продуктов и премий лучше задействовать с учетом возможного риска. Используя данные о местонахождении и сведения о страховых случаях с учетом погодных условий, система может определять категории риска на основании заявленных требований и затраченных сумм. Затем, используя модели, система будет оценивать новые заявления о страховой защите, классифицируя их по категории риска и возможному финансовому ущербу.
2. *Случайный лес (random forest)*. Данный алгоритм использует группу случайным образом созданных решающих деревьев с разными наборами атрибутов и дает им возможность проголосовать, чтобы выбрать самый популярный класс. Это универсальный, быстро обучаемый механизм для обнаружения связей внутри набора данных. В качестве примера можно привести нежелательные массовые рассылки, создающие повышенную

нагрузку на серверы. Автоматизированные методы фильтрации спама, разработанные для борьбы с этой проблемой, быстро и эффективно определяют нежелательные письма с помощью ансамбля решающих деревьев. Среди других применений стоит упомянуть диагностику заболеваний путем анализа медицинской карты пациента, распознавание банковских мошенничеств, прогнозирование числа звонков в колл-центрах и прогнозирование вероятности прибыли и убытка при покупке определенных акций.

3. *Метод ближайших соседей* (K-nearest neighbours, KNN). В этом достаточно эффективном алгоритме прогноз для нового объекта делается путём поиска K ближайших соседей в наборе данных и суммирования выходной переменной для этих K экземпляров. Сложность данного метода заключается в определении сходства между экземплярами данных. Если все признаки имеют один и тот же масштаб (например, сантиметры), то самый простой способ заключается в использовании евклидова расстояния – числа, которое можно рассчитать на основе различий с каждой входной переменной. KNN может потребовать много памяти для хранения всех данных, но зато быстро сделает прогноз. К тому же обучающие данные можно обновлять, чтобы прогнозы оставались точными с течением времени. Однако идея ближайших соседей плохо работает с многомерными данными, т.е. со множеством входных переменных, что негативно сказывается на эффективности алгоритма при решении задачи. Это явление получило название «проклятие размерности». Иными словами, в рамках этого алгоритма стоит использовать лишь наиболее важные для прогноза переменные.

4. *Наивный Байесовский классификатор*. Модель, построенная с помощью этого алгоритма, состоит из двух типов вероятностей, которые рассчитываются с помощью тренировочных данных: вероятность каждого класса и условная вероятность для каждого класса при каждом значении входной переменной. После расчёта вероятностной модели её можно использовать для прогноза по новым данным при помощи теоремы Байеса. Наивный Байес имеет такое

название, потому что алгоритм предполагает, что каждая входная переменная независимая. Это сильное предположение, которое не соответствует реальным данным. Тем не менее данный алгоритм весьма эффективен для целого ряда сложных задач вроде классификации спама или распознавания рукописных цифр.

5. *Нейронные сети.* Нейронные сети имитируют структуру головного мозга: каждый искусственный нейрон соединяется с несколькими другими нейронами. Нейросети имеют многослойную структуру: нейроны на одном слое передают данные нескольким нейронам на следующем и т.д. В конечном счете данные достигают выходного слоя, где сеть выдает предположение о том, как решить задачу, классифицировать объект и т.п. Нейросети применяются в целом ряде отраслей. В здравоохранении их используют при анализе медицинских снимков с целью ускорения диагностических процедур и поиска лекарств. В телекоммуникационной отрасли и медиаиндустрии нейросети можно применять для машинного перевода, распознавания мошенничеств и предоставления услуг виртуальных ассистентов. В финансовой отрасли их используют для распознавания мошенничеств, управления портфелями и анализа риска. В розничной торговле – для избавления от очередей и для персонализации обслуживания покупателей.
6. *Кластеризация.* Это группировка элементов данных, имеющих сходные характеристики, с помощью статистических алгоритмов. Это метод обучения без учителя, который можно использовать для решения задач классификации. Примером может быть сегментирование покупательской аудитории в зависимости от характеристик для уточнения адресации маркетинговых кампаний, рекомендации новостей конкретным читателям, помощь в работе правоохранительным органам. Кластеризация также действенна, когда в сложных наборах данных нужно обнаружить группы, которые трудно заметить без специальных средств: от группировки похожих документов в базе данных до обнаружения по криминальным новостям территорий с повышенным уровнем преступности.

7. *Поиск ассоциативных правил.* Это метод обучения без учителя, позволяющий находить зависимости между переменными. Широко используется в рекомендательных системах многих интернет-магазинов. Также ассоциативные правила позволяют выяснить, когда и при каких обстоятельствах покупатели приобретают те или иные сочетания товаров. Используя сведения о прошлых покупках и времени их совершения, можно составить программу скидок и сформировать индивидуальные предложения с расчетом на повышение продаж.

Таким образом, изучив ряд наиболее популярных на сегодняшний день методов и алгоритмов машинного обучения, а также особенности их применения в зависимости от решаемой задачи, можно сделать вывод, что алгоритмы машинного обучения могут успешно применяться в задачах построения моделей планирования показателей деятельности коммерческих банков.

Результаты и выводы первой главы

В первой главе проанализирована и структурирована информация о существующих подходах к планированию ключевых показателей эффективности деятельности кредитной организации, определены их преимущества и недостатки, а также современные области их активного применения. Кроме того, изучены современные методы машинного обучения, а также проведен обзор практики их применения в банковской сфере. По результатам исследования сделан вывод о применимости методов машинного обучения в задачах планирования показателей деятельности коммерческих банков.

2 ПОСТРОЕНИЕ РЕГРЕССИОННОЙ МОДЕЛИ ДЛЯ АНАЛИЗА КЛИЕНТСКОЙ БАЗЫ БАНКА

2.1 Построение полной бизнес-модели ПАО Сбербанк

ПАО Сбербанк (далее – Банк) – крупнейший банк в России, Центральной и Восточной Европе, один из ведущих международных финансовых институтов. Является лидером рынка по числу активных частных и корпоративных клиентов (свыше 100 млн клиентов), а также по числу активных ежемесячных пользователей в цифровых каналах (Сбербанк Онлайн и СМС-банк). Имеет сеть из 14 тыс. отделений и 70 тыс. устройств самообслуживания в России. Входит в топ-5 крупнейших работодателей России (286 тыс. сотрудников).

Миссия Банка – давать людям уверенность и надежность, делать их жизнь лучше, помогая реализовывать устремления и мечты [59].

У Банка есть четко сформулированная миссия, общие корпоративные цели и задачи (рисунок 1). Область, в которой анализируется выбранная для исследования проблема, заявлена в числе задач.



Рисунок 1 – Миссия, цели и задачи ПАО Сбербанк¹

На рисунке 2 представлена бизнес-стратегия Банка на 2021-2023 гг.

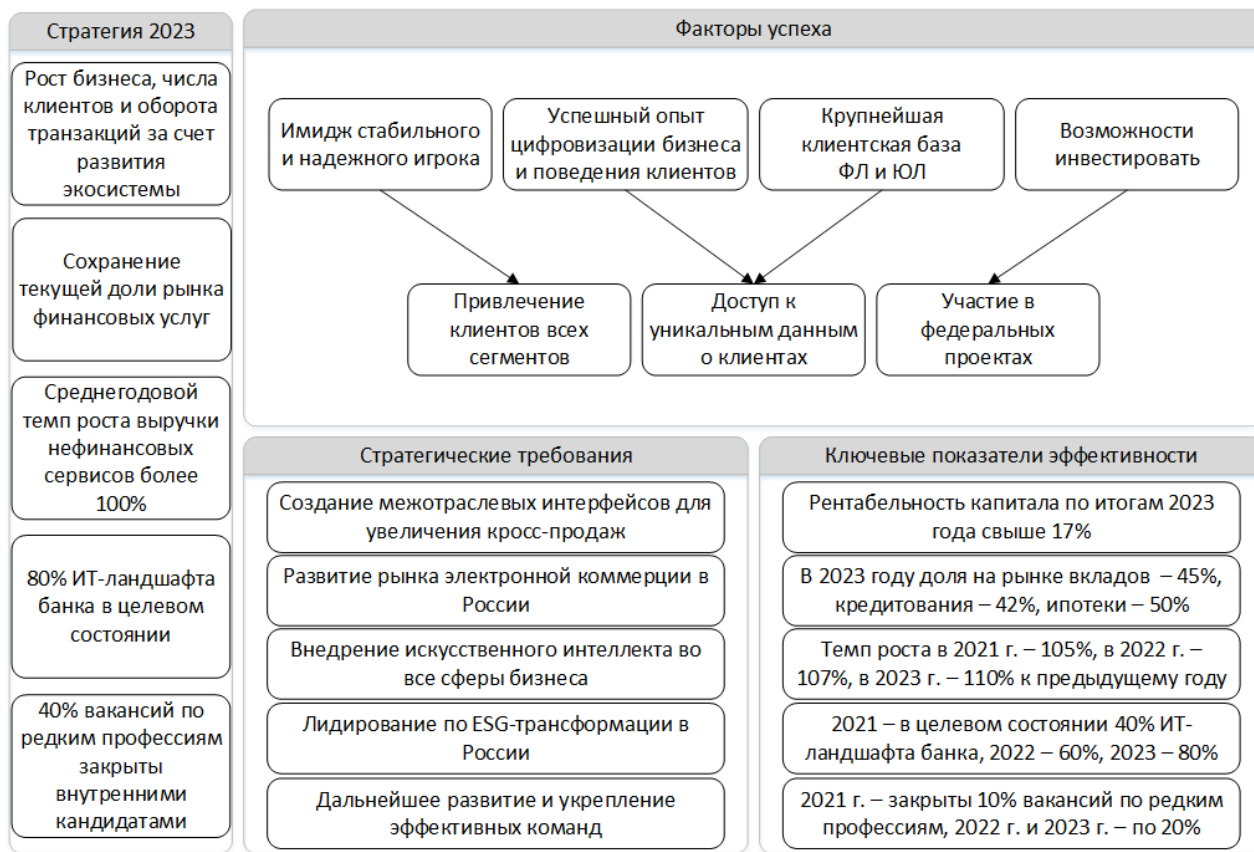


Рисунок 2 – Бизнес-стратегия ПАО Сбербанк на 2021-2023 гг.²

Таким образом, можно сделать вывод, что Банк имеет грамотную стратегию развития бизнеса в плане работы над ассортиментом и качеством своих продуктов и услуг, укрепления своих позиций на рынке, повышения квалификации персонала и управления всеми процессами.

В Банке выстроены все необходимые для успешной деятельности бизнес-процессы (рисунок 3). Однако необходимо их более подробное исследование для того, чтобы исключить проблемы внутри этих процессов.

¹ Составлено автором по: [61].

² Составлено автором по: [61, 60].

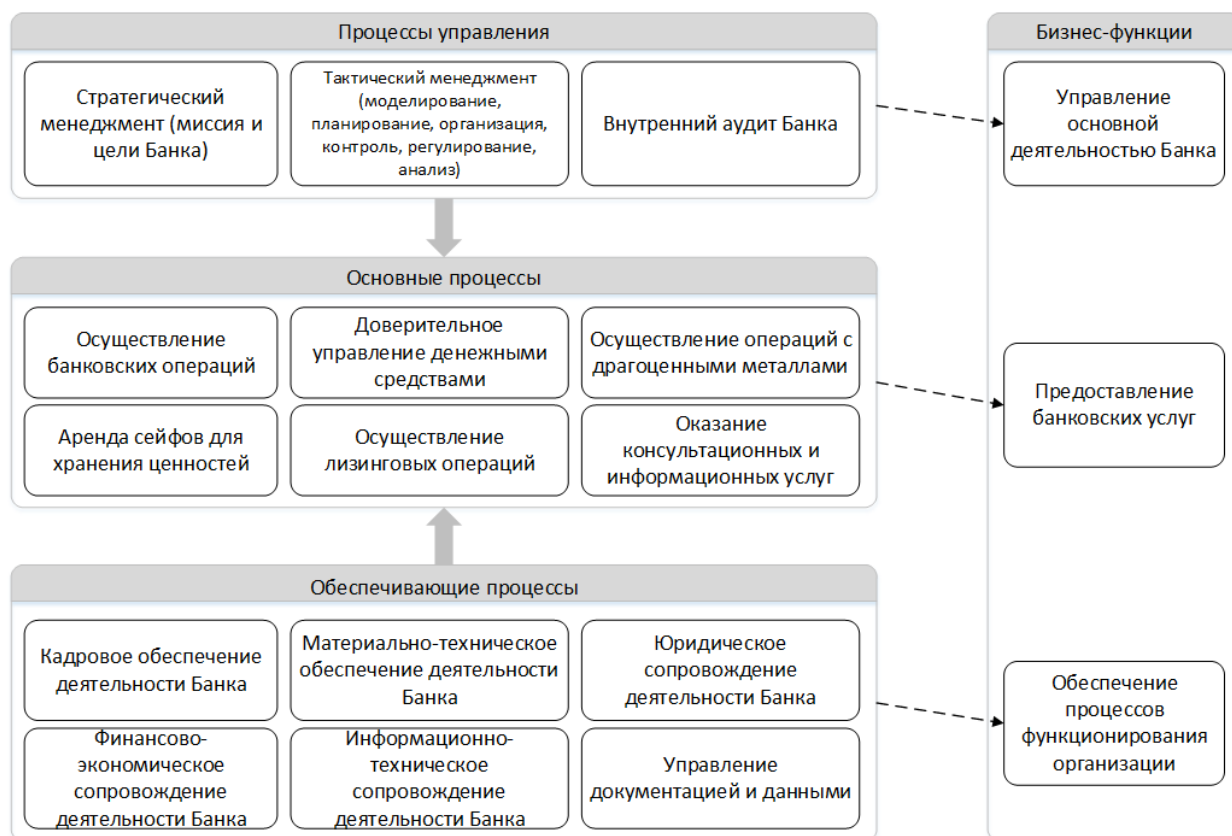


Рисунок 3 – Модель бизнес-процессов ПАО Сбербанк³

Бизнес-процессы требуют ресурсов для их исполнения. Лица, участвующие в ранее рассмотренных бизнес-процессах Банка представлены в матрице ответственности (таблица 1).

Таблица 1 – Матрица ответственности за бизнес-процессы Банка⁴

№ п/п	Процесс	Краткое описание	Ответственный за процесс, должность												
			Президент, Председатель Правления	Председатель территориального банка	Директор управления внутреннего аудита	Куратор бизнес-блока	Куратор блока «HR»	Куратор блока «Сервисы»	Куратор блока «GR, правовые вопросы, комплаенс и ДЗО»	Куратор блока «Финансы»	Куратор блока «Технологии»	Директор департамента управления делами			
1	Стратегический менеджмент	Долгосрочное планирование деятельности Банка	+												
2	Тактический менеджмент	Моделирование, планирование, организация, контроль, регулирование, анализ	+	+											

³ Составлено автором по: [61].

⁴ Составлено автором по: [61].

Окончание таблицы 1 – Матрица ответственности за бизнес-процессы Банка

№ п/п	Процесс	Краткое описание	Ответственный за процесс, должность											
			Президент, Председатель Правления	Председатель территориального банка	Директор управления внутреннего аудита	Куратор бизнес-блока	Куратор блока «HR»	Куратор блока «Сервисы»	Куратор блока «GR, правовые вопросы, комплаенс и ДЗО»	Куратор блока «Финансы»	Куратор блока «Технологии»	Директор департамента управления делами		
3	Внутренний аудит	Осуществление контроля за деятельностью Банка			+									
4	Основные процессы	Предоставление банковских услуг, доверительное управление, консультации и т.д.				+								
5	Кадровое обеспечение	Поиск и отбор кандидатов, прием на работу, увольнение, кадровое сопровождение					+							
6	Материально-техническое обеспечение	Поиск поставщиков, организация закупок, учет материальных ценностей						+						
7	Юридическое сопровождение	Договорная работа, юридическая экспертиза, представление интересов Банка в судах								+				
8	Финансово-экономическое сопровождение	Учет платежей, начисление заработной платы, подготовка отчетов для руководства									+			
9	Информационно-техническое сопровождение	Внедрение и поддержка автоматизированных систем, обеспечение хранения данных										+		
10	Управление документацией и данными	Обеспечение документооборота, регистрация, учет, рассылка, хранение и уничтожение документов												+

По данным таблицы 1 можно сделать вывод, что на каждый бизнес-процесс назначен ответственный, занимающий необходимую должность, поэтому все бизнес-процессы в Банке функционируют эффективно.

В Банке выстроена сложная организационная структура. В его состав входит 11 территориальных банков (далее – ТБ), подчиненных Центральному аппарату, отдельные подразделения центрального подчинения и ряд дочерних и зависимых организаций. ТБ, в свою очередь, имеют в своем подчинении отделения (84 на конец 2020 года). Отделениям подчиняются внутренние структурные подразделения (офисы), образующие широкую филиальную сеть по всей стране. На рисунке 4 представлена организационная структура Центрального аппарата. Нижестоящие подразделения имеют аналогичную структуру.

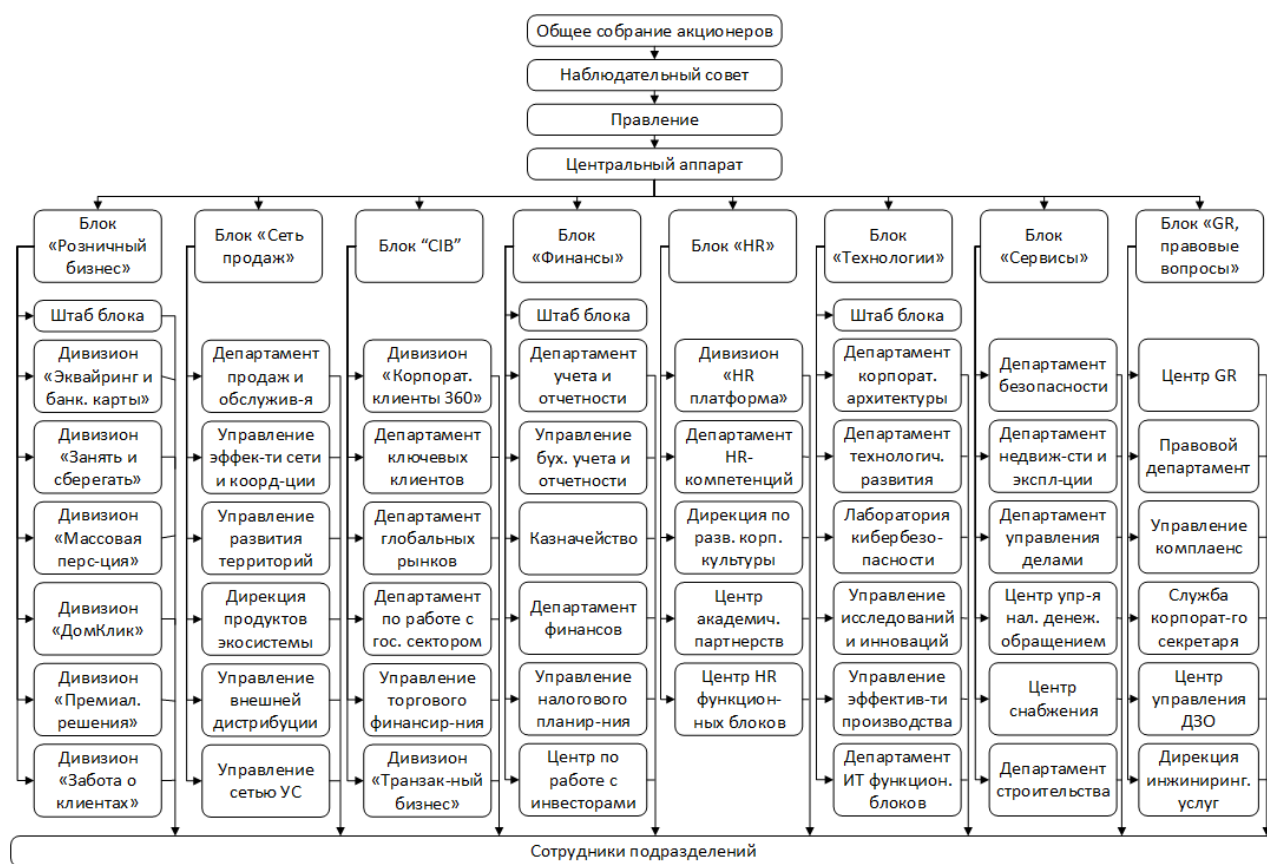


Рисунок 4 – Организационная структура Центрального аппарата Банка⁵

Таким образом, Банк имеет линейно-функциональную структуру управления. Оформленные по такому принципу организации, сохраняя простоту и жесткость линейных структур, приобретают специализированный управленческий потенциал. Выполнение управленческих функций на базе

⁵ Составлено автором по: [61, 1].

разграничения и специализации менеджмента обеспечивает рост качества управления всей организацией, повышение эффективности контроля линейных подразделений и достижение общеорганизационных задач [3, с. 62].

На рисунке 5 представлена модель продуктов и услуг Банка.

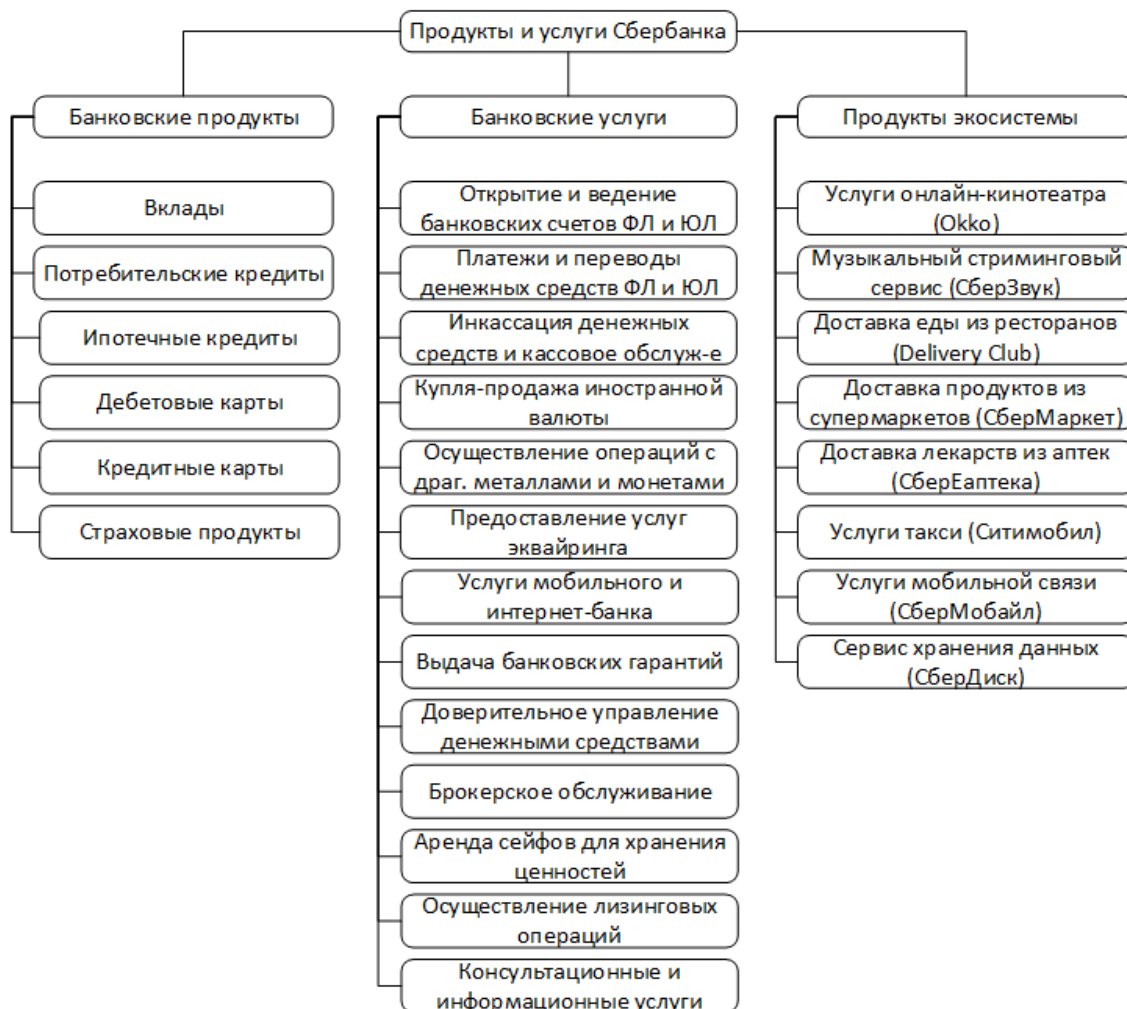


Рисунок 5 – Модель продуктов и услуг ПАО Сбербанк⁶

Анализ продуктов и услуг Банка показывает возможность расширения перечня предоставляемых услуг (например, за счет дальнейшего расширения экосистемы), что совпадает со стратегической целью организации, а значит руководство работает над этим вопросом.

⁶ Составлено автором по: [61, 2].

На рисунке 6 представлена модель связи бизнес-стратегии и бизнес-процессов Банка, из которой видно, что выстроенные в Банке бизнес-процессы эффективно взаимодействуют для достижения Стратегии 2023.

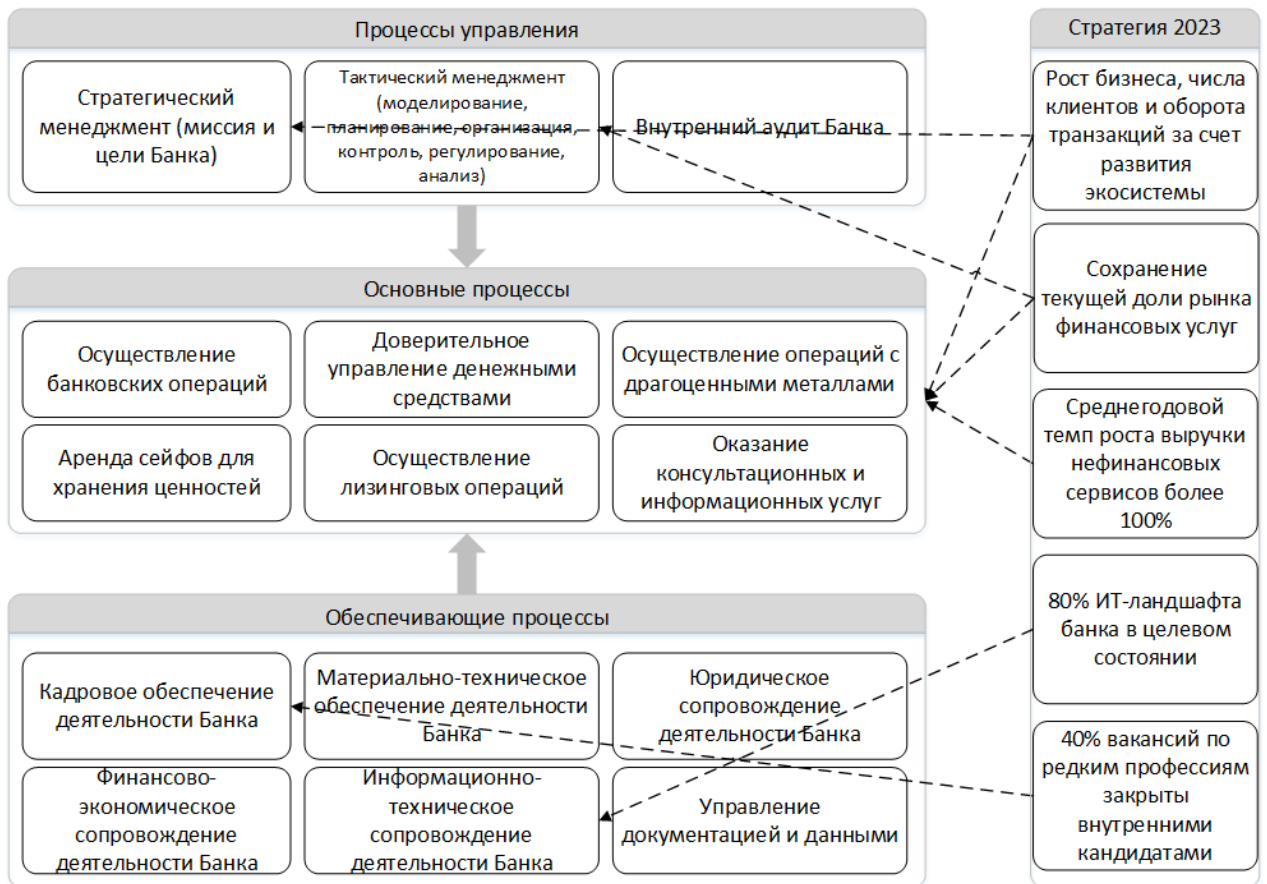


Рисунок 6 – Модель взаимосвязи бизнес-стратегии и бизнес-процессов Банка⁷

Данная взаимосвязь также иллюстрируется в матрице связей бизнес-стратегии и бизнес-процессов (таблица 2).

⁷ Составлено автором по: [61].

Таблица 2 – Матрица связей бизнес-стратегии и бизнес-процессов Банка⁸

№ п/п	Бизнес-процесс	Стратегические цели				
		Рост бизнеса, числа клиентов и оборота за счет развития экосистемы	Сохранение текущей доли рынка финансовых услуг	Среднегодовой темп роста выручки нефинансовых сервисов более 100%	80% ИТ-ландшафта банка в целевом состоянии	40% вакансий по редким профессиям закрыты внутренними кандидатами
1	Стратегический менеджмент	+++				
2	Тактический менеджмент		+++			
3	Внутренний аудит Банка					

Окончание таблицы 2 – Матрица связей бизнес-стратегии и бизнес-процессов Банка

№ п/п	Бизнес-процесс	Стратегические цели				
		Рост бизнеса, числа клиентов и оборота за счет развития экосистемы	Сохранение текущей доли рынка финансовых услуг	Среднегодовой темп роста выручки нефинансовых сервисов более 100%	80% ИТ-ландшафта банка в целевом состоянии	40% вакансий по редким профессиям закрыты внутренними кандидатами
4	Осуществление банковских операций	+++	+++	+++		
5	Доверительное управление	+++	+++	+++		
6	Операции с драгоценными металлами	+++	+++	+++		
7	Аренда сейфов для хранения ценностей	+++	+++	+++		
8	Осуществление лизинговых операций	+++	+++	+++		
9	Оказание консульт. и информац. услуг	+++	+++	+++		
10	Кадровое обеспечение					+++
11	Материально-техническое обеспечение					
12	Юридическое сопровождение					
13	Финансово-экономич. сопровождение					
14	Информац.-технич. сопровождение				+++	
15	Управление документацией и данными					

В результате анализа все цели были сопоставлены с соответствующими бизнес-процессами, что обеспечивает возможность их достижения и говорит о грамотной организации работы Банка.

Для автоматизации бизнес-процессов любая современная организация использует различные информационные системы и технологии. В Банке функционирует ряд автоматизированных систем собственной разработки, а

⁸ Составлено автором по: [61].

также программное обеспечение сторонних поставщиков. Собственные системы постоянно модернизируются, стороннее ПО регулярно обновляется до актуальных версий. Схема основных информационных систем и сервисов, используемых в Банке, представлена на рисунке 7.

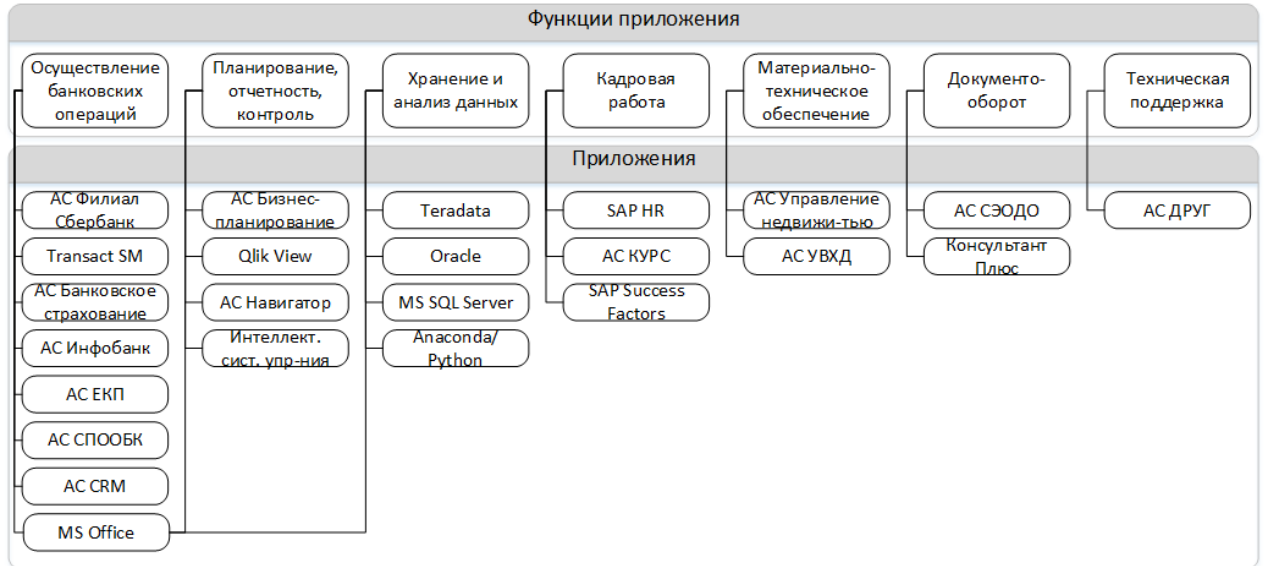
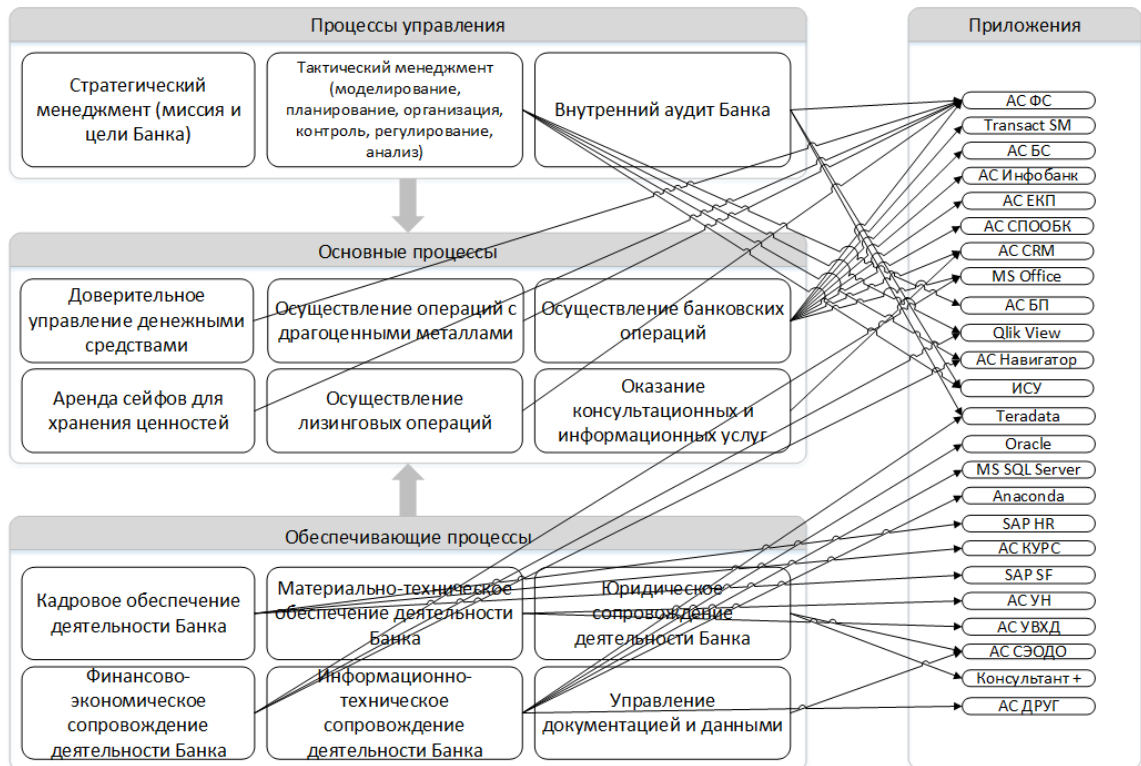


Рисунок 7 – Схема информационных систем и сервисов Банка⁹

Связь бизнес-процессов с используемыми в Банке приложениями приведена на рисунке 8. По данным схемы можно сделать вывод, что все важные



⁹ Составлено автором по: [61].

процессы автоматизированы и используют актуальное программное обеспечение.

Рисунок 8 – Схема связи бизнес-процессов и приложений Банка¹⁰

В таблице 3 представлена матрица использования приложений специалистами.

Таблица 3 – Матрица использования приложений специалистами¹¹

№	Участники использования	Описание использования	Приложения																							
			АС ФС	Transact SM	АС БС	АС Инфобанк	АС ЕКП	АС СПООбК	АС CRM	MS Office	АС БП	Qlik View	АС Навигатор	ИСУ	Teradata	Oracle	MS SQL Server	Anaconda	SAP HR	АС КУРС	SAP SF	АС УН	АС УВХД	АС СЭОДО	Консультант +	АС ДРУГ
1	Руководство	Оперативное управление и контроль								+		+	+	+												
2	Аудиторы	Внутренний контроль деятельности	+							+				+	+											
3	Сотрудники филиальной сети	Продажа продуктов, обслуживание клиентов	+	+	+	+	+	+	+				+													
4	Сотрудники подразделения планирования и отчетности	Анализ данных, разработка моделей планирования, подготовка отчетов								+	+	+		+				+								
5	Специалисты блока HR	Кадровое сопровождение деятельности								+									+	+	+					
6	Специалисты службы МТО	Материально-техническое обеспечение								+												+	+			
7	Специалисты юридической службы	Юридическое сопровождение деятельности								+														+	+	
8	Специалисты блока "Финансы"	Финансово-экономическое сопровождение								+		+	+													
9	Специалисты блока "Технологии"	Информационно-техническое сопровождение								+				+	+	+	+									+
10	Сотрудники управления делами	Документооборот								+														+		

¹⁰ Составлено автором по: [61].

¹¹ Составлено автором по: [61].

Результаты анализа показывают, что у всех специалистов имеется необходимое для работы программное обеспечение актуальных версий.

На рисунке 9 представлена общая схема ИТ-инфраструктуры Банка, обеспечивающая предоставление информационных, вычислительных и телекоммуникационных ресурсов, возможностей и услуг сотрудникам Банка.

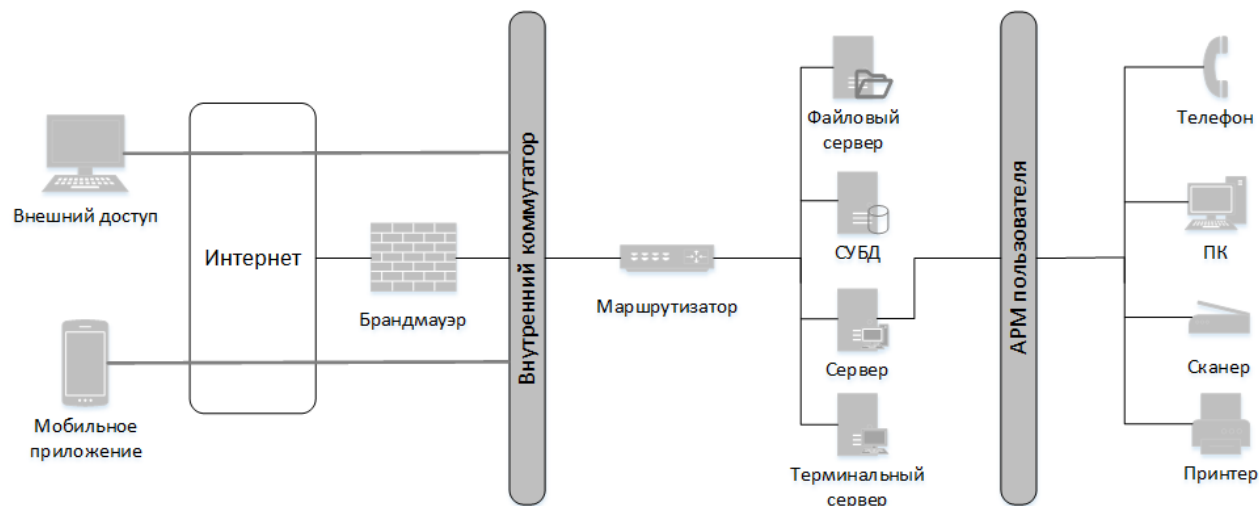


Рисунок 9 – Схема ИТ-инфраструктуры Банка¹²

Полная модель архитектуры Банка представлена на рисунке 10.

¹² Составлено автором по: [61].

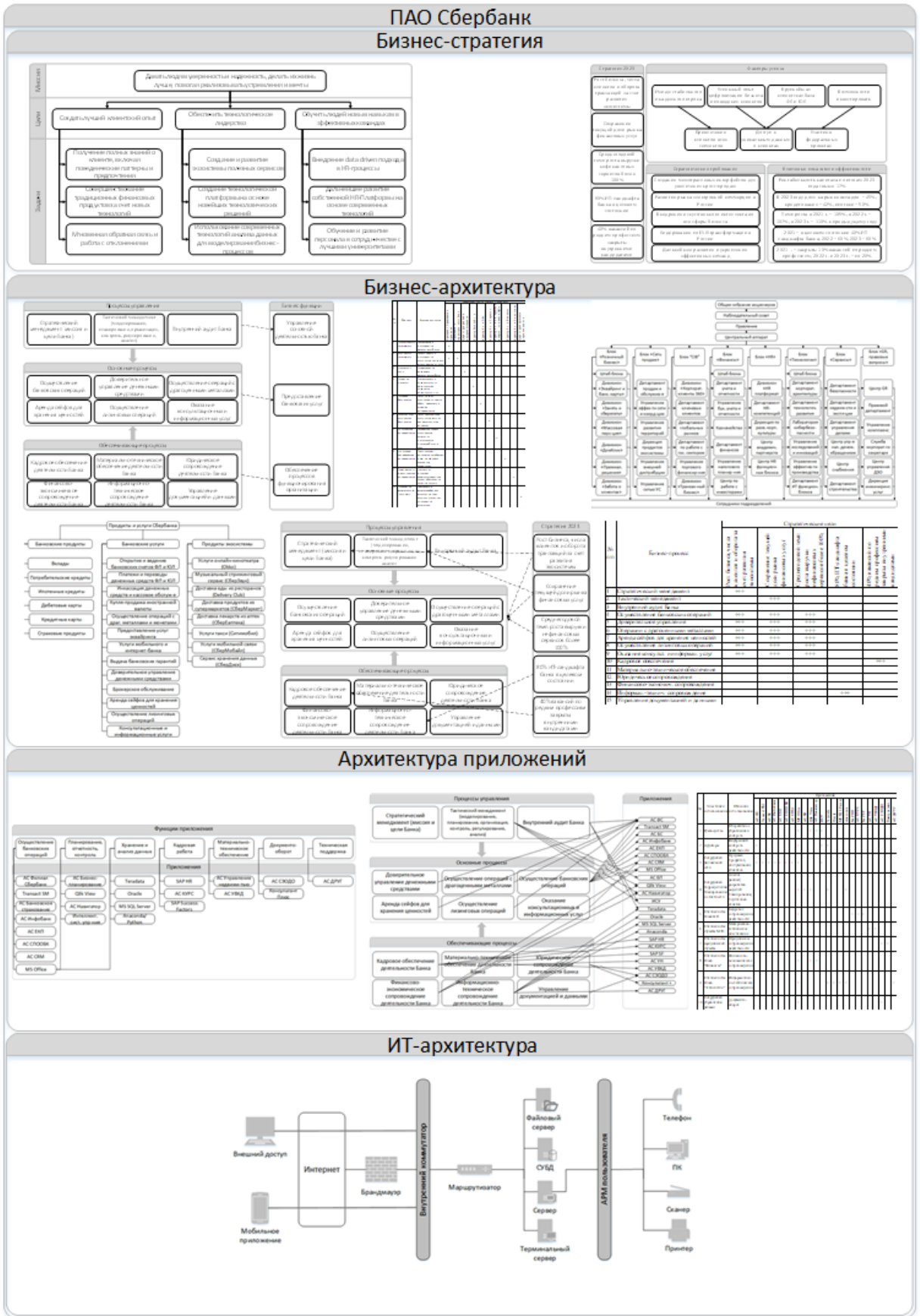


Рисунок 10 – Полная модель бизнес-архитектуры ПАО Сбербанк¹³

¹³ Составлено автором по: [61].

2.2 Системный анализ проблемы

В Розничном бизнесе Банка вся клиентская база разделена на 3 клиентских сегмента (массовый, высокодоходный и премиальный), каждый из которых обслуживается в отдельном канале. Так, клиенты высокодоходного сегмента обслуживаются в канале «Сбербанк Премьер». У этого канала есть свои КРІ, по которым оцениваются результаты его работы. Одним из показателей является величина комиссионного дохода (далее – КД).

Планы по КД на предстоящий период рассчитываются с помощью модели планирования, разработку и поддержку которой осуществляет подразделение планирования. В настоящее время руководство канала «Сбербанк Премьер» считает неудовлетворительным качество модели, поскольку она недостаточно хорошо отражает реальную ситуацию в бизнесе, из-за чего продающие подразделения не могут выстроить оптимальную стратегию продаж на своих территориях и обеспечить выполнение планов.

Во время очередной плановой кампании на 2021 год на совещании руководителей канала, подразделения планирования и представителей руководства Банка проблема была рассмотрена подробнее, результаты представлены в таблице 4.

Таблица 4 – Первоначальная постановка проблемы Банка¹⁴

Элемент	Описание
Проблема	низкое качество модели планирования КД
воздействует на	продающие подразделения канала «Сбербанк Премьер»
результатом чего является	сложность построения оптимальной стратегии продаж

Все заинтересованные лица согласились с наличием данной проблемы и определили, что ее решение сможет улучшить существующие процессы, в связи с чем возникла необходимость в проведении более глубокого анализа выбранной проблемы с целью выявления путей ее решения.

¹⁴ Составлено автором по: [28].

Для анализа в рамках данного исследования выбрана подсистема планирования. Ее целью является расчет планов по КД клиентского сегмента «Сбербанк Премьер» на предстоящий период, а в задачи входит разработка алгоритма модели, реализация модели, доведение планов до подразделений, коммуникации по вопросам планирования с заинтересованными лицами.

Участниками процесса выступают сотрудники подразделения планирования Центрального аппарата, сотрудники управления планирования и отчетности ТБ, руководство канала «Сбербанк Премьер».

Модель бизнес-процесса AS-IS представлена на рисунке 11.



Условные обозначения:
КД - Комиссионный доход;
ТБ - Территориальные банки;
ЦА - Центральный аппарат;
ЦПО - Центр персонального обслуживания;

Рисунок 11 – Модель AS-IS процесса «Планирование КД СБ Премьер»¹⁵

¹⁵ Составлено автором по: [39].

На рисунках 12–16 представлена декомпозиция указанного процесса.

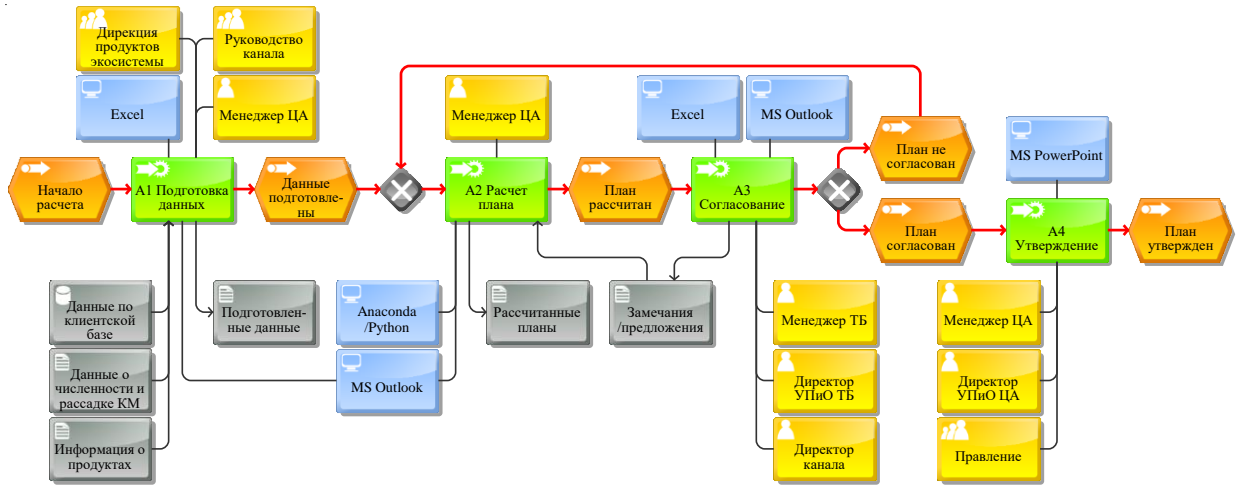


Рисунок 12 – Общая схема процесса «Планирование КД СБ Премьер»¹⁶

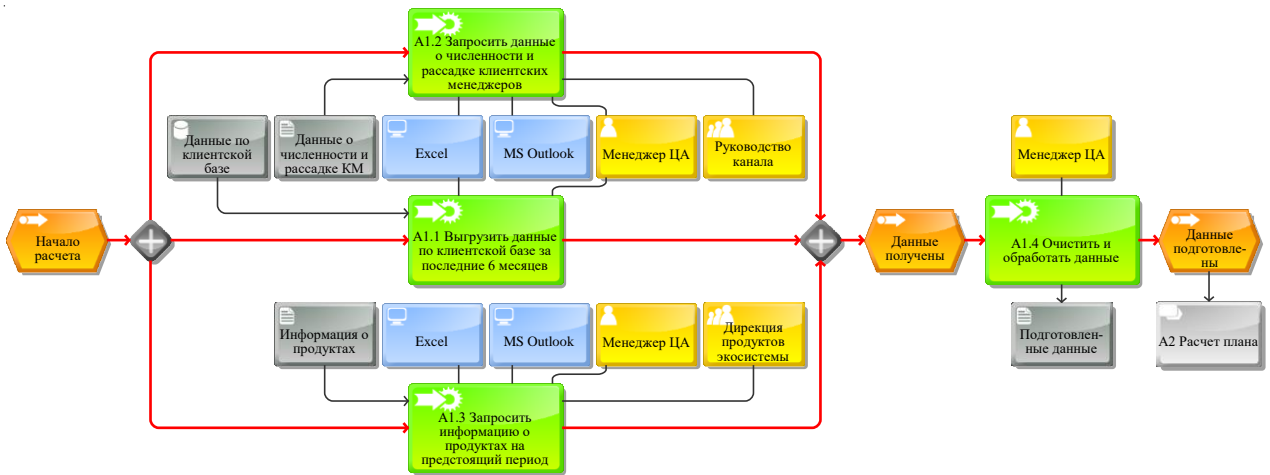


Рисунок 13 – Схема процесса «A1 Подготовка данных»¹⁷

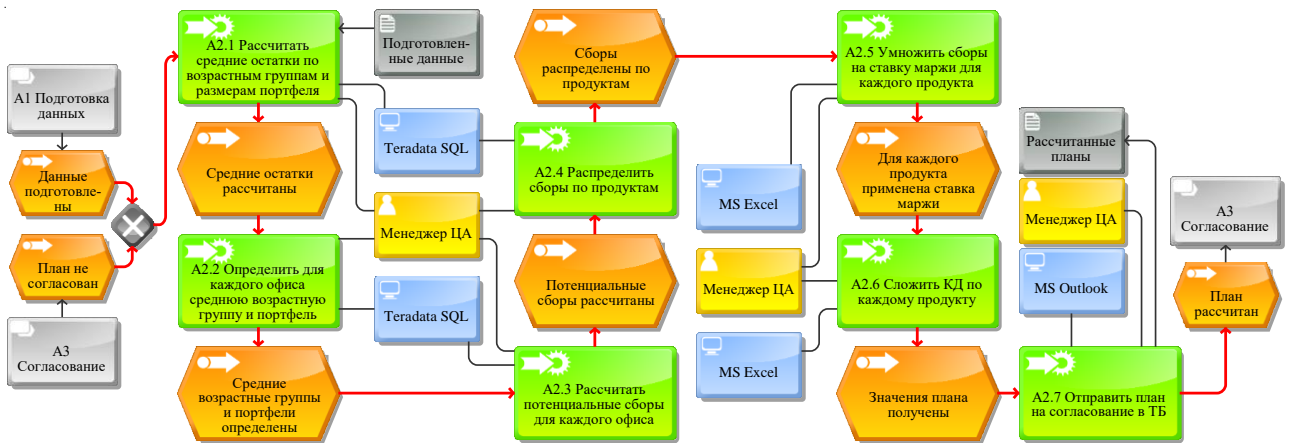


Рисунок 14 – Схема процесса «A2 Расчет плана»¹⁸

¹⁶ Составлено автором по: [39].

¹⁷ Составлено автором по: [39].

¹⁸ Составлено автором по: [39].

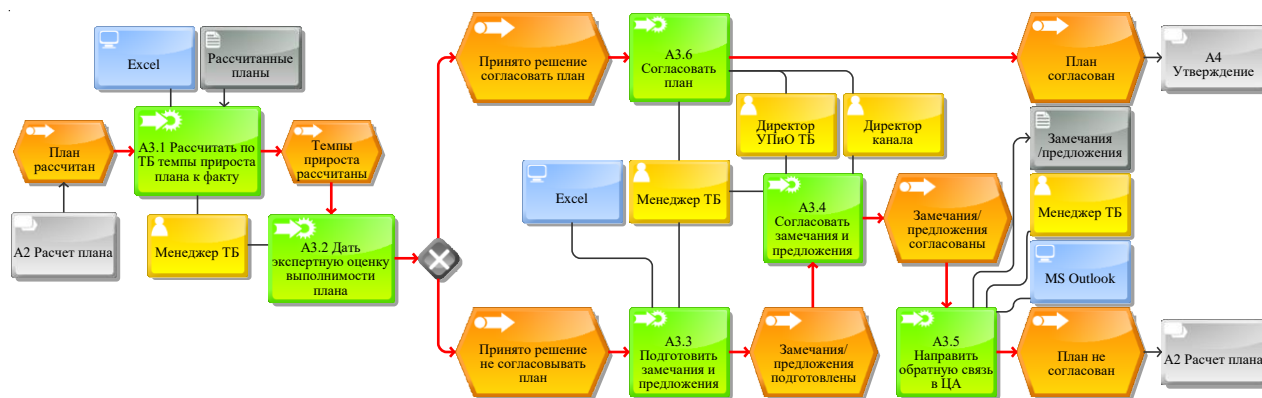


Рисунок 15 – Схема процесса «А3 Согласование»¹⁹



Рисунок 16 – Схема процесса «А4 Утверждение»²⁰

Для глубинного понимания исследуемой проблемы необходимо провести ее системный анализ.

Анализ проблемы – это процесс осознания реальных проблем и потребностей пользователя и предложения решений для удовлетворения этих потребностей [28, с. 59].

Для рассмотрения реальной проблемы и ее причин можно использовать различные методы [26, с. 98], [25, с. 40], [47, с. 31], [40, с. 56]:

- деревья проблем;
- диаграмма Исикавы (причинно-следственная диаграмма «рыбий скелет»);
- профиль причин;
- диаграмма связей и др.

¹⁹ Составлено автором по: [39].

²⁰ Составлено автором по: [39].

В целях исследования проблемы низкого качества модели планирования КД, для ее изучения была использована технология построения дерева проблем [26, с. 97–98], состоящая из следующих шагов:

1. Формулировка проблемы.

Условие этапа: участники построения дерева проблем должны быть компетентны в исследуемой области, рассматриваемая проблема должна существовать в настоящем, а не в прошлом или будущем, формулировка должна быть конкретной, без лишних слов.

Таким образом, была составлена формулировка проблемы: низкое качество планов по КД.

2. Выявление заинтересованных лиц.

Условие этапа: необходимо выявить всех *заинтересованных лиц* – участников, которых прямо или косвенно касается данная проблема, и установить, каким именно образом то или иное заинтересованное лицо зависит от проблемы. Для этого нужно ответить на следующие вопросы:

- а) на кого эта проблема оказывает самое большое воздействие?
- б) кто будет непосредственно участвовать в решении проблемы?
- в) какие организации или группы людей могут оказать влияние на ход работы?

Для рассматриваемой проблемы Банка были выявлены следующие заинтересованные лица:

- а) руководители канала «Сбербанк Премьер»;
- б) клиентские менеджеры;
- в) менеджеры подразделений планирования и отчетности ЦА и ТБ.

3. Начало построения дерева проблем.

Условие этапа: дерево проблем должно быть построено при участии прямых заинтересованных лиц и экспертов, разбирающихся в рассматриваемом вопросе, а также состоять из трех частей: корней, ствола и кроны. *Корни* – это причины, из-за которых возникла проблема. Именно они обуславливают ее существование. Если их устранить, проблема исчезнет. *Ствол* – формулировка

проблемы. *Крона* – это любые последствия, которые повлекла за собой проблема. На данном этапе был сформирован ствол (рисунок 17).

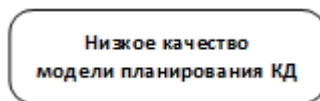


Рисунок 17 – Построение ствола дерева проблем²¹

4. Выявление корней.

Условие этапа: участниками должно быть определено максимальное число причин возникновения проблемы, так как именно их решение окажет решающее воздействие. Затем производится их группировка с указанием взаимосвязей. В данном исследовании заинтересованные лица Банка выявили и сгруппировали возможные корни. Результат представлен на рисунке 18.

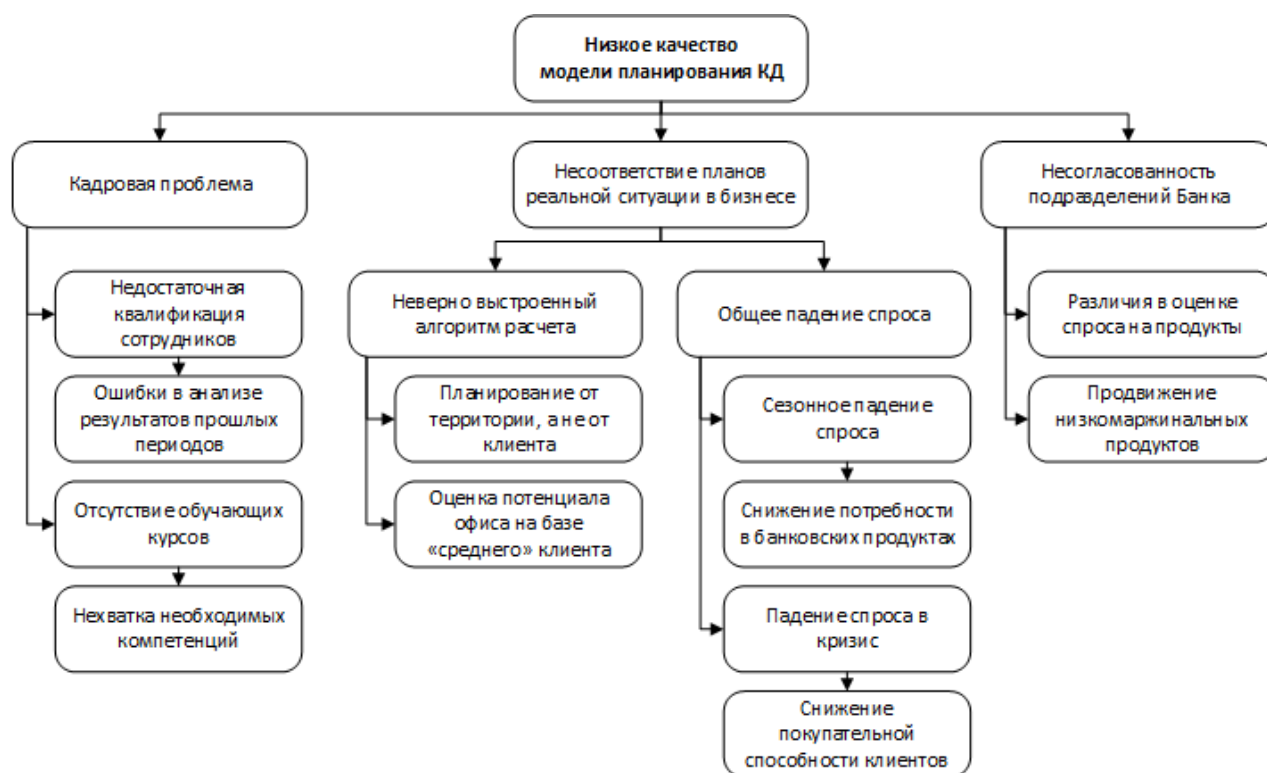


Рисунок 18 – Результат построения корней дерева проблем²²

²¹ Составлено автором по: [26].

²² Составлено автором по: [26].

5. Завершение. Построение кроны.

Условие этапа: участниками должны быть определены последствия каждой выявленной проблемы. При этом декомпозиция проводится до тех пор, пока последствия остаются в рамках проблемы.

В данном исследовании заинтересованные лица Банка выявили и сгруппировали возможную крону, завершив дерево проблем (рисунок 19).

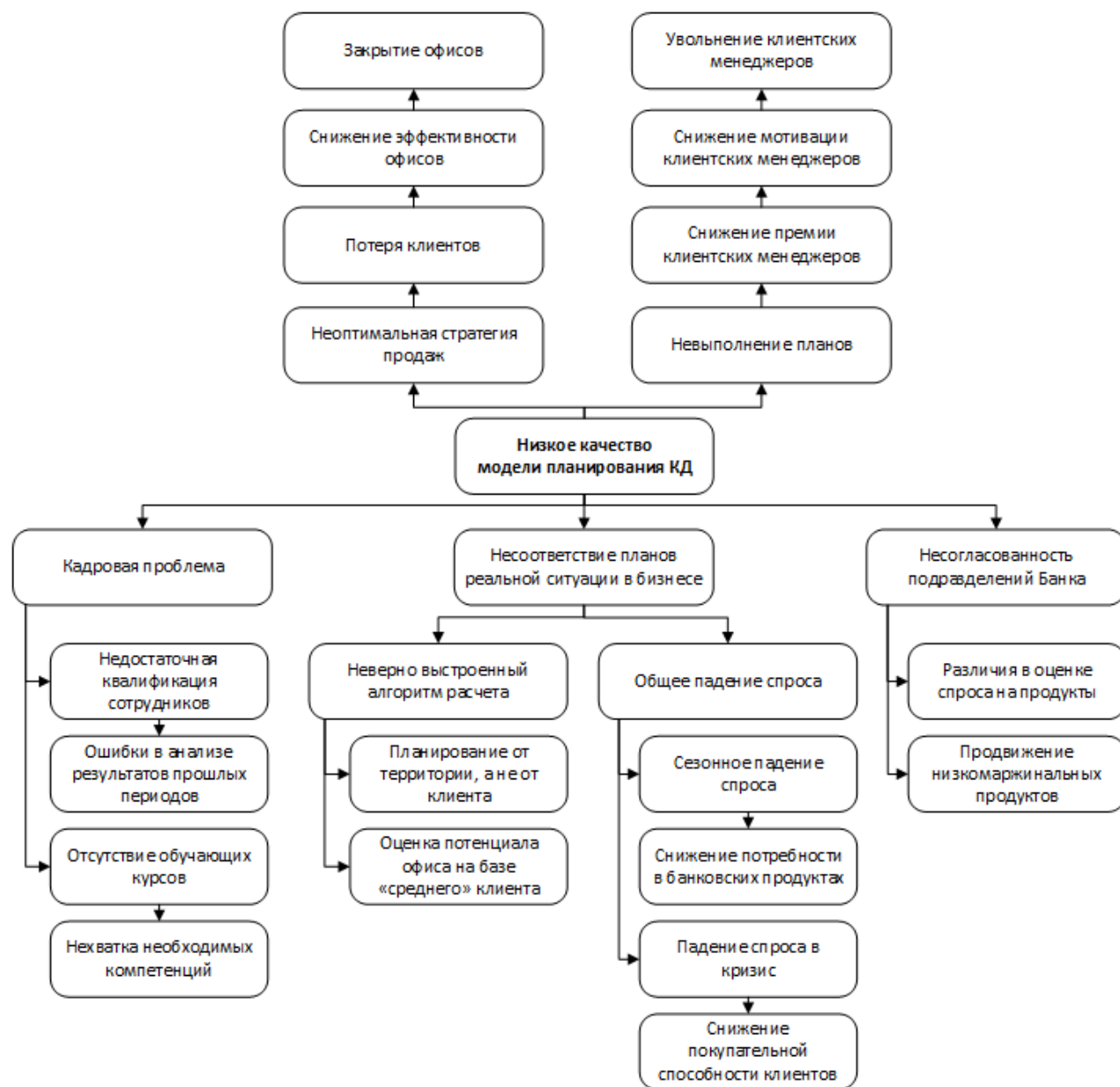


Рисунок 19 – Результат построения кроны дерева проблем²³

²³ Составлено автором по: [26].

6. Анализ результатов. Постановка задачи решения проблемы.

Условие этапа: на данном этапе участники процесса исследования проблемы должны изучить полученное дерево проблем, проверить возможные причины возникновения проблемы, отсеять неактуальные, и расставить приоритеты решения.

Заинтересованные лица Банка провели анализ собственной деятельности на наличие выявленных причин низкого качества модели планирования КД.

Исследования показало, что HR-служба Банка эффективно взаимодействует с подразделениями планирования и отчетности, уровень профессиональных компетенций менеджеров по планированию был оценен как достаточный для выполнения стоящих перед подразделениями задач. Кроме того, Банк предоставляет сотрудникам широкие возможности для обучения и развития необходимых навыков. Это значит, что кадровая проблема на данный момент не актуальна для решения, однако Банк постоянно держит ее на контроле, как одну из приоритетных для существования всей организации.

Также не подтвердилось и общее падение спроса на банковские продукты. По данным отчетности Банка за 2019 г., фактический КД канала «Сбербанк Премьер» возрастал в среднем на 10% в квартал. В 2020 г. из-за кризиса, вызванного пандемией коронавируса, произошло существенное снижение КД во 2 квартале – на 25% по сравнению с 1 кварталом 2020 г. и на 12% по сравнению со 2 кварталом 2019 г. Для нивелирования негативных последствий кризиса было принято решение во 2 квартале 2020 г. премировать клиентских менеджеров и руководителей канала по нормативному коэффициенту 1 независимо от выполнения плана, а при планировании на 2 полугодие 2020 г. применить скидки к планам. Таким образом, проблема падения спроса либо не актуальна, либо оперативно решается в процессе осуществления деятельности.

Анализ порядка взаимодействия подразделений Банка в процессе планирования не подтвердил гипотезу о несогласованности между ними. Различия в оценке будущего спроса на продукты между дирекцией продуктов экосистемы и подразделением планирования, как правило, удается

урегулировать в ходе плановой кампании с участием руководства канала и, при необходимости, вышестоящего руководства. Поэтому к моменту утверждения планов подразделения приходят к компромиссной оценке рынка. Активное продвижение продуктов с низкой маржинальностью, например, в целях расширения клиентской базы или продажи имиджевого продукта, не оказывает существенного влияния на качество модели планирования, поскольку данный вопрос также прорабатывается заинтересованными подразделениями и проходит урегулирование в процессе планирования.

Наиболее вероятной причиной низкого качества модели планирования был определен неверно выстроенный алгоритм расчета. В основе модели лежит определение потенциальных сборов каждого офиса на базе усредненного клиента данного офиса по возрастной группе и размеру портфеля под управлением. Такой подход приводит к несоответствию планов реальной ситуации в бизнесе, поскольку клиенты, которые обслуживаются в офисе, различаются по своей возрастной группе и достатку и даже в разных офисах в пределах одного города наблюдается различный интерес клиентов к тем или иным банковским продуктам. Таким образом, на повестку был вынесен вопрос о поиске иного подхода к планированию КД с целью большего соответствия модели бизнесу и повышения качества планов.

Проведенный анализ позволил отсеять не актуальные на данный момент проблемы, в результате чего осталась одна приоритетная ветвь, требующая срочного решения (рисунок 20).

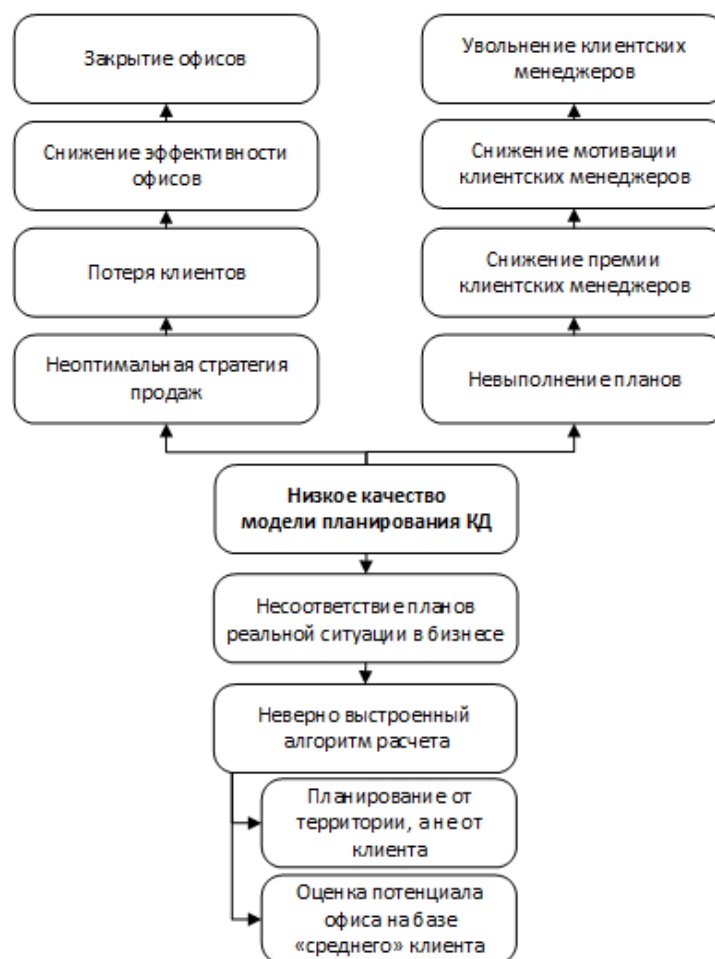


Рисунок 20 – Дерево проблем Банка после проведения анализа²⁴

Дерево проблем наглядно показывает причины возникновения проблемы и очевидно, что невозможно устранить основную проблему, не устранив причины ее возникновения. Чтобы устранить причины, необходимо понять, каким должен быть результат той работы, которая приведет к устранению одной из проблем, то есть определить цели системного анализа.

Определить цель системного анализа означает ответить на вопрос, что надо сделать для снятия проблемы. *Сформулировать цель* – значит указать направление, в котором следует двигаться, чтобы разрешить существующую проблему, показать пути, которые уводят от существующей проблемной ситуации.

²⁴ Составлено автором по: [26].

Для каждой проблемы есть решение, и исходя из описания проблемы, можно его сформулировать. Для того чтобы преобразовать проблемы в цели, каждой негативной ситуации надо противопоставить положительную ситуацию, в которой причина проблемы устранена. Таким образом, будет получено *дерево целей* – иерархический перечень (структура подчиненных элементов) последовательно детализированных целей, построение которого напоминает древовидную структуру – цели более низкого уровня являются разъяснением (детализацией) способа достижения целей верхнего уровня.

В данном исследовании проблемы Банка были преобразованы в цели, в результате чего было получено дерево целей и следствий, представленное на рисунке 21.

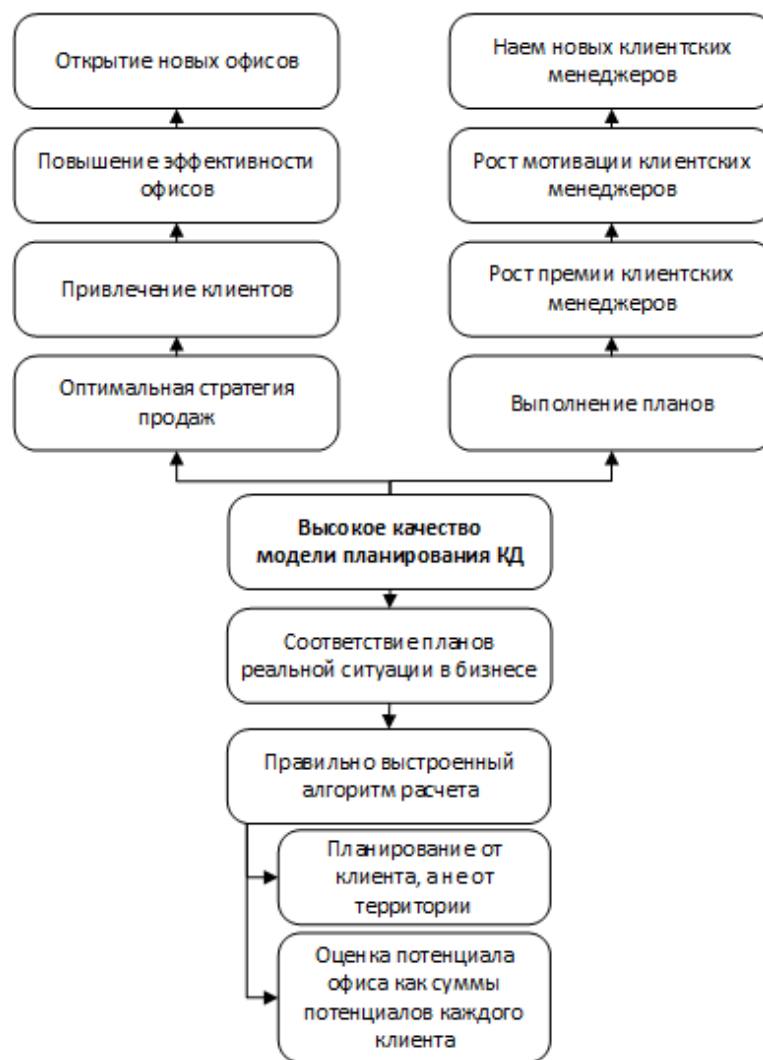


Рисунок 21 – Дерево целей и следствий решения проблемы Банка²⁵

²⁵ Составлено автором по: [26].

Изучив полученные результаты, рабочая группа, занимавшаяся исследованием проблемы Банка, составила план дальнейшей работы: изменить подход к построению модели планирования КД, а именно производить оценку потенциала офиса как суммы потенциалов каждого клиента, поскольку существующий подход не обеспечивает соответствия модели реальной ситуации в бизнесе и может привести к увольнению клиентских менеджеров, падению эффективности офисов и их закрытию.

Таким образом, была дополнена и конкретизирована первоначальная постановка проблемы (таблица 5).

Таблица 5 – Завершение постановки проблемы Банка²⁶

Элемент	Описание
Проблема	низкого качества модели планирования КД
воздействует на	продающие подразделения канала «Сбербанк Премьер»
результатом чего является	сложность построения оптимальной стратегии продаж и невыполнение планов, что может привести к увольнению сотрудников и закрытию офисов
Выигрыш от	изменения подхода к построению модели планирования
может состоять в	разработке менеджерами оптимальной стратегии продаж и обеспечении выполнения планов, что может привести к открытию новых офисов и привлечению новых клиентов

После содержательной формулировки целей необходимо задать критерии и ограничения при которых будет осуществляться поиск возможных вариантов их достижения. Данный этап необходимо проводить каждый раз, когда цели проекта будут конкретизироваться, дополняться или разбиваться на подцели и задачи, так как каждой новой цели и задаче соответствуют свои критерии и ограничения для достижения.

Критерий (от греч. *kriterion* – средство для суждения) – это признак, на основании которого производится оценка, определение или классификация чего-либо; мерило, оценка [16, с. 66]. Критерии так же, как и цели, могут быть представлены в виде *дерева критериев*.

²⁶ Составлено автором по: [28].

В качестве критерия достижения целей часто выбирают *показатель эффективности системы*, который может выражаться как в качественной, так и в количественной форме.

При разработке показателей эффективности необходимо придерживаться следующих правил [42, с. 146]:

- 1) набор показателей должен содержать минимально необходимое их количество для обеспечения полноценного управления бизнес-процессом;
- 2) каждый показатель должен быть измерим;
- 3) стоимость измерения показателя не должна превышать управленческий эффект от использования данного показателя.

Наряду с заданными критериями большое влияние на выбор того или иного варианта решения оказывает система выделенных *ограничений* – условий, отражающих влияние внешних и внутренних факторов, которые нужно учитывать в задаче принятия решений.

При этом, необходимо учитывать, что существуют различные источники ограничений:

- организационные;
- экономические;
- правовые;
- технические;
- экологические;
- эксплуатационные;
- психологические и т. д.

Качественные ограничения формулируются, как правило, в терминах «не разрешается», «не допускается», а количественные – «не более», «не менее», «в интервале от–до». Ограничения, как правило, дополняют (конкретизируют) сформулированные ранее цели и в ряде случаев могут сделать цели нереализуемыми. В этом случае необходимо через проведение ряда итерационных процедур снять часть ограничений.

При формировании целей, критериев и ограничений используется так называемое *пространство целеполагания* – совокупность систем, предъявляющих требования к исследуемой системе. В пространство целеполагания также включается сама система, предъявляющая собственные требования [41, с. 20].

В данном исследовании критерий достижения главной цели повышения качества модели планирования КД был сформулирован следующим образом: не менее 90% офисов «Сбербанк Премьер» выполняют план на полугодие на 100% и более.

К задаче был применен ряд ограничений:

- поиск решения в течении не более чем двух месяцев с момента начала анализа проблемы;
- реализация решения в как можно более быстрые сроки, но не позднее начала плановой кампании на 2 полугодие 2021 г. (май-июнь 2021 г.);
- бюджет проекта равен размеру нормативной оплаты труда менеджеров подразделения планирования, занятых в разработке и реализации проекта;
- допускается привлечение на время проекта сторонних специалистов и ресурсов, если необходимость этого обоснована;
- согласование с руководством и документирование всех этапов проекта.

Задача изменения подхода к построению модели планирования КД, являясь подцелью главной цели, также требовала определения собственных критериев и ограничений. Критерием достижения данной подцели была определена реализация подхода к оценке потенциала офиса как суммы потенциалов всех клиентов данного офиса для дальнейшей работы по достижению главной цели – повышения качества модели планирования КД. Ограничения представлены на рисунке 22.

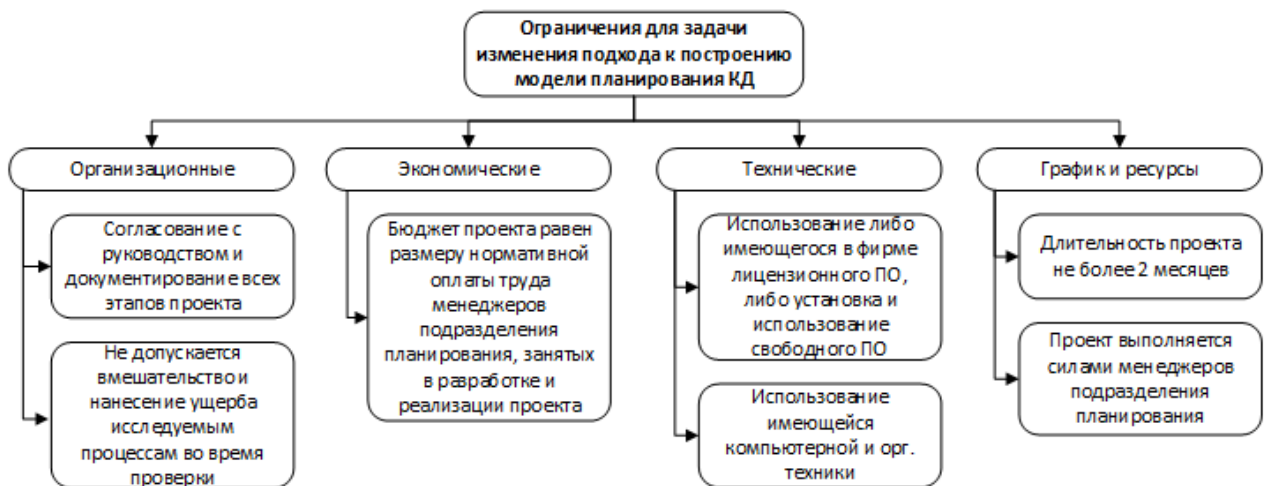


Рисунок 22 – Дерево ограничений для изменения подхода к построению модели планирования КД²⁷

Полученные на данном этапе формулировки целей и задач решения проблемы, а также критерии и ограничения к выполнению были задокументированы и переданы на исполнение назначенным ответственным лицам Банка.

Бизнес-требования соответствуют критериям решения главной проблемы. Необходимо лишь трансформировать цели и критерии проекта в бизнес-требования.

Таким образом, будущее решение должно обеспечить выполнение плана на полугодие не менее 90% офисов «Сбербанк Премьер» на 100% и более. Это и есть высокоуровневая цель проекта – бизнес-требование.

²⁷ Составлено автором по: [41].

2.3 Разработка алгоритма построения регрессионной модели на основе деревьев решений

Перед началом анализа данных необходимо получить представление о банковских продуктах, предлагаемых клиентам «Сбербанк Премьер» и о том, какие из них приносят Банку комиссионный доход. Модель продуктов «Сбербанк Премьер» представлена на рисунке 23.

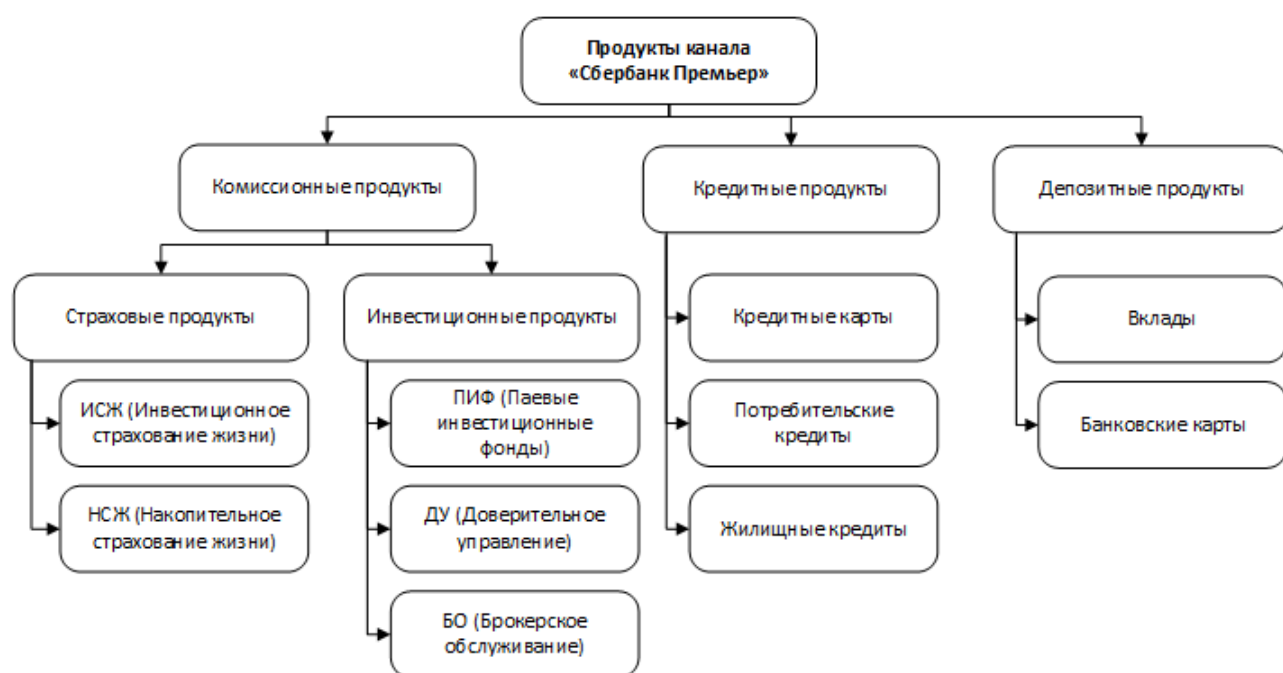


Рисунок 23 – Модель банковских продуктов «Сбербанк Премьер»²⁸

К основным этапам интеллектуального анализа данных относятся [36, с. 453]:

1. Извлечение, преобразование и загрузка данных в хранилище данных.
2. Хранение данных и управление ими в многомерных базах данных.
3. Предоставление доступа к данным бизнес-аналитикам с помощью прикладного программного обеспечения.
4. Представление проанализированных данных в легко понятных формах, таких как графики.

²⁸ Составлено автором по: [61].

Данные о клиентах высокодоходного сегмента «Сбербанк Премьер» содержатся в централизованном хранилище данных на базе СУБД Teradata. Их обработкой, загрузкой и хранением занимаются ответственные подразделения Центрального аппарата.

Витрина данных содержат большое количество информации в разрезе клиентов, включая возраст и пол, количество и объем операций покупок по категориям, количество и объем поступлений и списаний денежных средств по источникам и получателям, наличие кредитных продуктов, остатки по депозитам и дебетовым картам, объем денежных средств, размещенных в инвестиционных и страховых продуктах и т.д. – всего более 250 полей.

В целях реализации проекта по изменению подхода к модели планирования комиссионного дохода были выгружены данные по клиентским портфелям – объемам денежных средств, размещенных на депозитах и банковских картах, а также в инвестиционных и страховых продуктах – за период январь–июль 2020 г.

В первую очередь, была изучена структура продуктов клиентов «Сбербанк Премьер» по возрастным группам (рисунок 24).

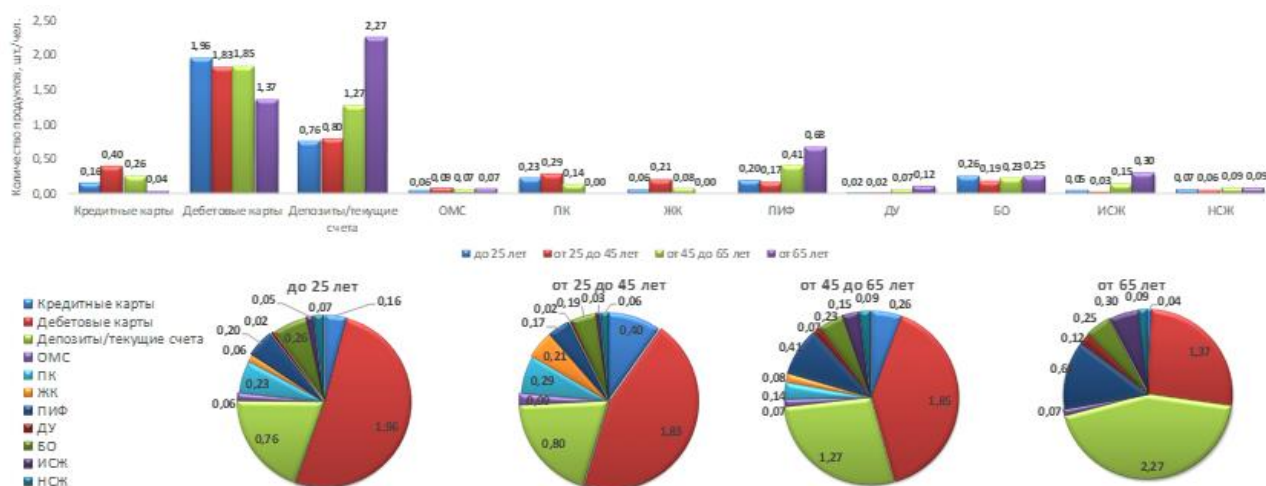


Рисунок 24 – Структура продуктов клиентов «Сбербанк Премьер» по возрастным группам²⁹

²⁹ Составлено автором по: [36].

Далее был рассчитан текущий уровень проникновения продуктов, (рисунок 25) который показывает наиболее востребованные продукты. Так, например, среди инвестиционных инструментов наиболее популярны брокерские счета – они открыты у 16,8% клиентов.

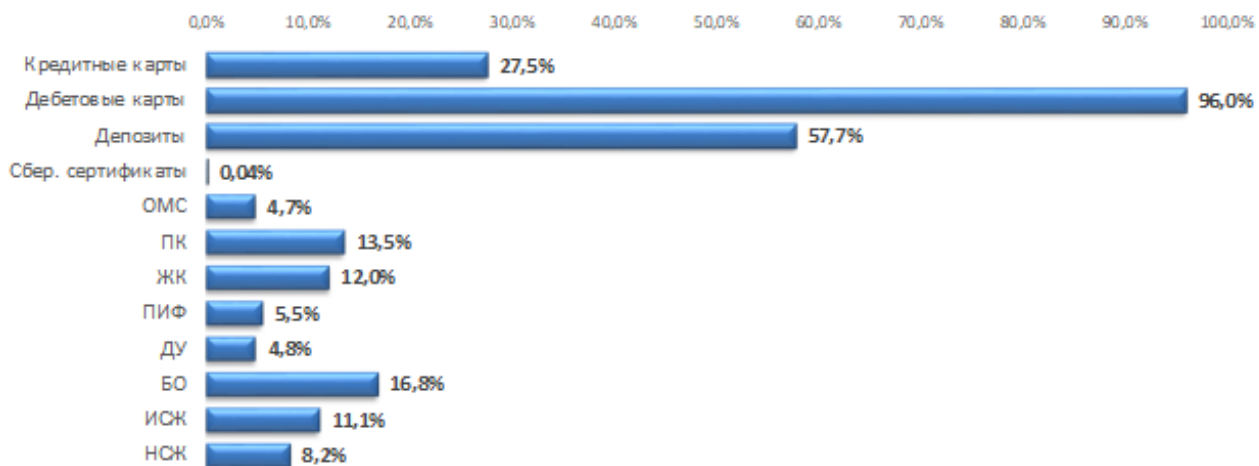


Рисунок 25 – Уровень проникновения продуктов «Сбербанк Премьер»³⁰

На следующем шаге клиенты были сгруппированы по возрастам и по размеру активов, а затем внутри каждой группы были рассчитаны доли каждого продукта в структуре портфеля.

По всем выборкам было рассчитано стандартное отклонение, которое показало, что распределение долей внутри группы достаточно сильно разбросано относительно среднего значения. Поэтому для оценки наиболее характерной картины по группе была выбрана медиана – срединное значение доли в выборке. Результат анализа представлен на рисунке 26.

³⁰ Составлено автором по: [36].

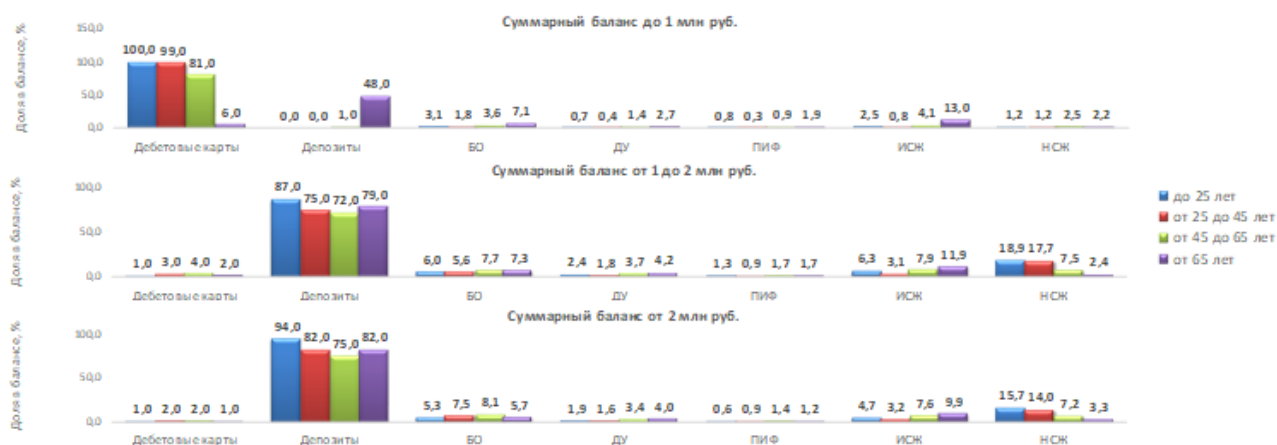


Рисунок 26 – Медианные доли продуктов в структуре баланса по размеру активов в разрезе возрастных групп³¹

Анализ диаграммы позволил сделать следующие выводы:

- клиенты с суммарным балансом менее 1 млн руб. предпочитают держать все свои средства на дебетовых картах;
- при портфеле от 1 млн руб. средства, главным образом, размещаются во вклады (медианные доли по всем возрастным группам выше 70%). Также возрастает интерес к страховым продуктам, в частности, к НСЖ (накопительное страхование жизни);
- при анализе структуры портфеля выделились четко выраженные группы: нулевые доли (группа клиентов, которые не держат средства в данном продукте), 100%-е доли (группа клиентов, чьи средства в полном объеме размещены в каком-либо одном продукте) и промежуточные доли.

На следующем шаге были определены клиенты со 100%-ми долями того или иного продукта и рассмотрена их динамика по месяцам (рисунок 27).

³¹ Составлено автором по: [36].

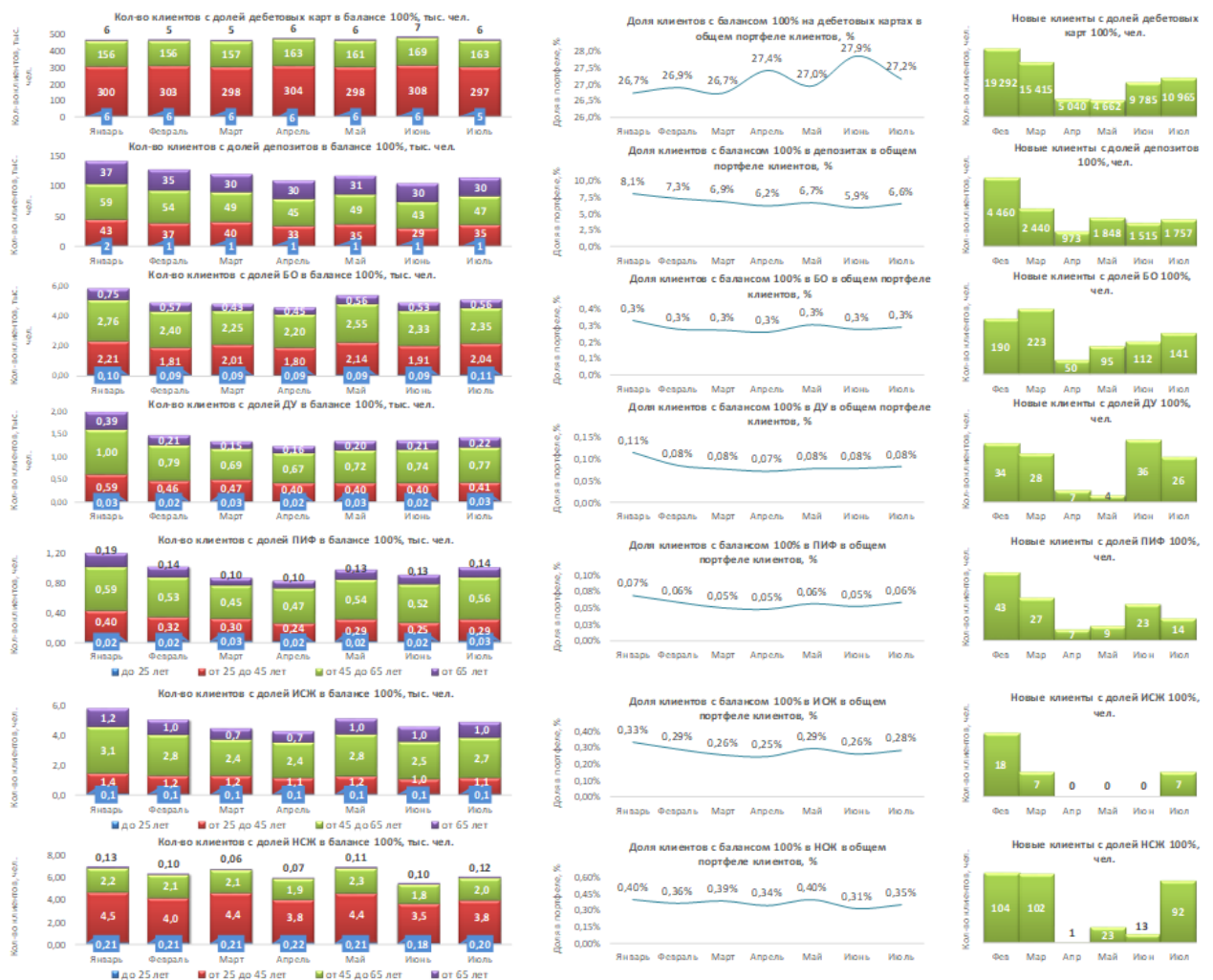


Рисунок 27 – Анализ клиентов со 100%-ной долей продукта³²

Данный анализ показал, что суммарно около 33% клиентов держат все свои средства на счетах дебетовых карт либо в депозитах. Эти средства можно рассматривать как свободные деньги для размещения.

Результат дальнейшего анализа клиентов, у которых 100% средств лежат на дебетовых картах, представлен на рисунке 28.

³² Составлено автором по: [36].



Рисунок 28 – Анализ клиентов со 100%-ной долей продукта³³

По данным анализа были сделаны следующие выводы:

- 76% рассмотренных клиентов имеют остаток менее 100 тыс. руб. Как правило, такие клиенты не склонны к инвестициям и используют эти деньги на текущие расходы;
- для оставшихся 24% (117 тыс. клиентов) с остатками более 100 тыс. руб. могут быть сформированы потенциально интересные инвестиционные предложения.

Проведенный анализ подтверждает изначальную гипотезу о целесообразности внедрения в модель планирования комиссионного дохода подхода «от клиента», то есть имеет смысл рассчитывать склонность каждого клиента к покупке того или иного продукта, используя деревья решений. Анализ ретроданных по клиентской базе «Сбербанк Премьер» показал, что такие критерии, как возрастная группа, размер активов и доля продукта в суммарном балансе, могут быть значимыми признаками при построении деревьев решений.

Планирование комиссионного дохода «Сбербанк Премьер» производится раз в полугодие в рамках общей плановой кампании Банка. Затем каждый месяц расчет корректируется с учетом входящей базы клиентов и их остатков. Существующий алгоритм расчета планов выглядит следующим образом:

1. Вся клиентская база распределяется по группам в зависимости от возраста и суммарного размера портфеля под управлением, который включает в себя все

³³ Составлено автором по: [36].

деPOSITные и комиссионные продукты клиента. Затем для каждой группы рассчитывается суммарный средний остаток по комиссионным продуктам за последнее полугодие. Расчет производится с помощью скрипта SQL с использованием методов математической статистики.

2. Далее для каждого офиса по данным за последнее полугодие определяется средняя возрастная группа клиентов и средний размер суммарного портфеля. Для этого используется отдельный скрипт SQL.
3. Следующий скрипт SQL на основании данных, полученных на первых двух этапах, определяет потенциальный объем сборов, т.е. средств, размещенных одним клиентом в комиссионные продукты. Умножение этого значения на среднюю величину клиентской базы позволяет получить суммарный потенциальный объем сборов для каждого офиса.
4. На следующем этапе потенциальные сборы офиса распределяются по отдельным комиссионным продуктам в зависимости от доли каждого продукта в данной группе клиентов. Для этого применяется еще один скрипт SQL.
5. Дальнейшие расчеты производятся в MS Excel. Объем сборов по каждому продукту умножается на средневзвешенную ставку комиссионного дохода для данного продукта за последнее полугодие. Таким образом определяется величина комиссионного дохода по каждому продукту.
6. Итоговый план по комиссионному доходу для каждого офиса рассчитывается путем суммирования комиссионных доходов по всем продуктам.

Подход к планированию комиссионного дохода «Сбербанк Премьер», предлагаемый в рамках данной работы, основывается на методах машинного обучения. Модель планирования будет реализована на языке программирования Python. Для данного языка разработаны специальные пакеты с различными статистическими и экономическими инструментами. Пакеты представляют собой библиотеки для работы специфических функций или специальных областей применения. В составе библиотек программирования существуют как

стандартные статистические инструменты, например, линейная регрессия, так и современные (бэггинг, случайный лес, градиентный бустинг и др.).

Анализ данных по клиентской базе, проведенный ранее в данном параграфе, показал, что доля конкретного комиссионного продукта в портфеле клиента является значимой метрикой. Поэтому именно доля продукта будет использоваться в качестве целевой переменной. Поскольку ее значения не являются классами, а определяются некоторой функцией, то в модели для ее прогнозирования целесообразно применять метод случайного леса (random forest), который представляет собой ансамбль регрессионных деревьев решений. В пользу выбора именно этого метода говорит также и то, что для прогнозирования целевой переменной будут использоваться многомерные данные со множеством входных переменных, что может являться проблемой для других алгоритмов, например, для метода ближайших соседей (KNN).

Предлагаемый алгоритм расчета планов по комиссионному доходу «Сбербанк Премьер» выглядит следующим образом:

1. По данным за последнее полугодие для каждого клиента рассчитывается медиана доли каждого комиссионного продукта в суммарном портфеле. Дополнительно рассчитываются такие метрики, как медианный объем портфеля, медианный остаток на депозитах, банковских картах, в инвестиционных и страховых продуктах, а также медианная доля продукта в соответствующей группе продуктов (доля ИСЖ и НСЖ в страховом портфеле, доля ПИФ, ДУ и БО в инвестиционном портфеле). Это исходные данные, которые рассчитываются с помощью скрипта SQL.
2. Полученные на первом этапе данные загружаются в датафрейм из библиотеки Pandas в Python. Затем для каждого комиссионного продукта строится предиктивная модель с помощью объекта RandomForestRegressor из библиотеки Sklearn. Целевой функцией в каждой модели является доля соответствующего продукта в суммарном портфеле клиента, а прогноз ее значения производится по набору входных переменных, которые являются

значимыми признаками для данного продукта. Методика выбора значимых признаков для каждого продукта рассмотрена далее в данной главе.

3. С помощью отдельного скрипта SQL готовится выборка данных по каждому клиенту, содержащая значения существенных признаков на начало текущего полугодия, а также информацию об офисе обслуживания.
 4. Полученные на предыдущем этапе данные загружаются в отдельный датафрейм.
 5. Для каждого комиссионного продукта в соответствующую модель загружаются значения существенных признаков и рассчитывается прогноз доли данного продукта в суммарном портфеле клиента.
 6. Полученные значения долей умножаются на суммарный портфель клиента, и определяются плановые объемы средств клиентов, размещенных в каждом комиссионном продукте.
 7. Объемы средств умножаются на средневзвешенные ставки по комиссионным продуктам, и таким образом рассчитывается плановый комиссионный доход по каждому продукту.
 8. Комиссионные доходы по каждому продукту суммируются в итоговый план по комиссионному доходу для каждого клиента.
 9. Полученные планы по комиссионному доходу переносятся в результирующий датафрейм, содержащий данные об офисе обслуживания, и производится суммирование планов с группировкой по территориальным банкам, головным отделениям и офисам. Полученный датафрейм выгружается в MS Excel.
- Предложенный алгоритм необходимо встроить в существующий бизнес-процесс. Модель ТО-ВЕ бизнес-процесса «Планирование КД СБ Премьер» представлен на рисунке 29.



Условные обозначения:
 КД - Комиссионный доход;
 ТБ - Территориальные банки;
 ЦА - Центральный аппарат

Рисунок 29 – Модель TO-BE процесса «Планирование КД СБ Премьер»³⁴

³⁴ Составлено автором по: [39].

Предлагаемые изменения влияют только на подпроцесс «A2 Расчет плана». Декомпозиция данного процесса ТО-ВЕ представлена на рисунке 30.

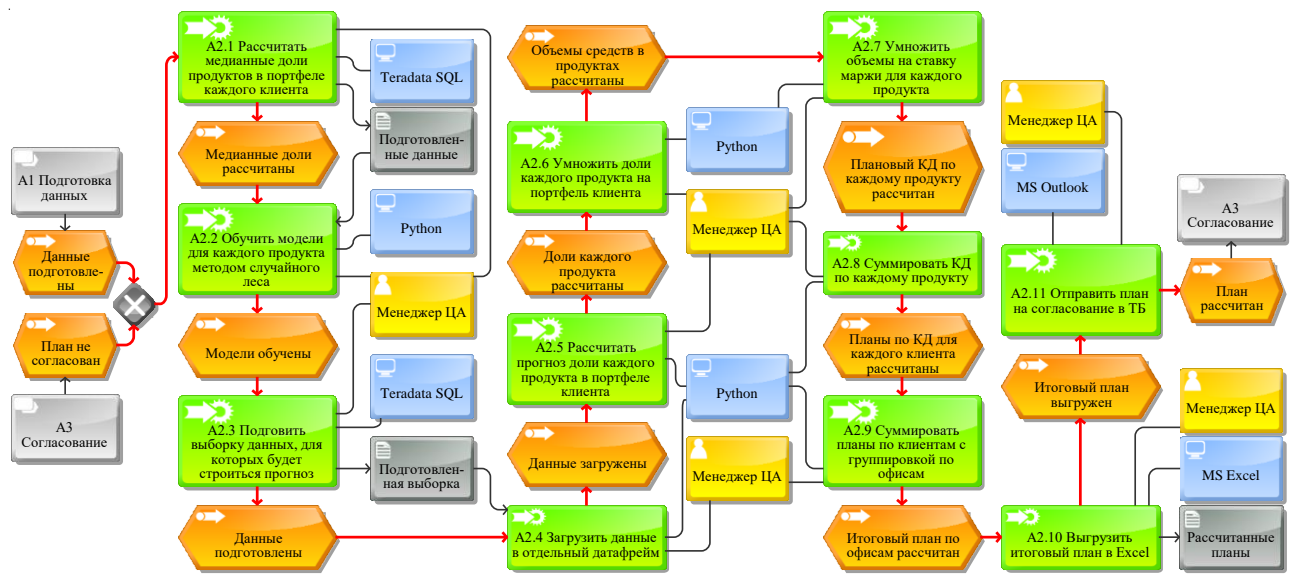


Рисунок 30 – Схема ТО-ВЕ процесса «A2 Расчет плана»³⁵

Результаты и выводы второй главы

Во второй главе построена полная модель бизнес-архитектуры ПАО Сбербанк, построена модель бизнес-процесса планирования комиссионного дохода «Сбербанк Премьер» AS-IS, определены основные недостатки существующего процесса, проведен системный анализ проблемы, по результатам которого сделан вывод о необходимости изменения текущего подхода к модели планирования. Кроме того, построена модель процесса ТО-ВЕ, а также разработана методика реализации проекта по совершенствованию существующего процесса планирования.

³⁵ Составлено автором по: [39].

3 РАЗРАБОТКА МОДЕЛИ ПЛАНИРОВАНИЯ КОМИССИОННОГО ДОХОДА КАНАЛА «СБЕРБАНК ПРЕМЬЕР»

3.1 Постановка задачи и описание источника данных

Центральной проблемой существующего бизнес-процесса планирования комиссионного дохода клиентского сегмента «Сбербанк Премьер» является неудовлетворительное качество модели планирования. Системный анализ данной проблемы, проведенный в предыдущей главе (п. 2.2), показал, что корневой причиной низкого качества планов является текущий подход к построению модели планирования. В его основе лежит принцип определения для каждого офиса среднестатистического клиента. По данным о клиентской базе рассчитывается ряд усредненных характеристик клиентов, которые обслуживаются в данном офисе, в частности, средняя возрастная группа и средний портфель денежных средств клиента. Затем на основе статистики по всей клиентской базе сегмента «Сбербанк Премьер» определяется сплит продуктов, которые необходимо продавать данному офису в течение планового периода, после чего полученная продуктовая разбивка переводится в плановый комиссионный доход.

Данный подход приводит к получению плановых значений, которые недостаточно хорошо отражают реальную ситуацию в бизнесе. Клиенты, которые обслуживаются в конкретном офисе, различаются по своей возрастной группе и достатку и даже в разных офисах в пределах одного города наблюдается различный интерес клиентов к тем или иным банковским продуктам. Поэтому те банковские продукты, которые предлагает модель для продажи в данном офисе, зачастую могут просто не пользоваться спросом у клиентов. Отсюда у руководства офиса возникают сложности с построением оптимальной стратегии продаж и, как следствие, проблемы с выполнением планов.

По итогам системного анализа проблемы был поставлен вопрос об изменении подхода к планированию комиссионного дохода.

Задача сформулирована следующим образом: необходимо реализовать модель планирования комиссионного дохода клиентского сегмента «Сбербанк Премьер».

Ключевое бизнес-требование: модель должна обеспечить выполнение плана на полугодие не менее 90% офисов «Сбербанк Премьер» на 100% и более.

В основе предлагаемого подхода лежит анализ каждого клиента на предмет его склонности к размещению средств в те или иные банковские продукты. Предлагается обучить алгоритм на данных о структуре портфеля денежных средств клиента в разные периоды времени и с помощью построенной модели спрогнозировать, в какие продукты будут размещены его средства в течение будущего периода.

Предлагаемый подход будет способствовать более тщательному отбору продуктов для продажи в каждом офисе и таким образом лучше отражать реальную ситуацию в бизнесе, что в результате приведет к росту продаж и в конечном итоге к выполнению планов.

Основные требования и ограничения:

- горизонт планирования – полугодие;
- проверка соответствия модели ключевому бизнес-требованию должна быть произведена на фактических значениях объема комиссионного дохода «Сбербанк Премьер» за 2 полугодие 2020 года в разрезе офисов и на плановых значениях, рассчитанных моделью на тот же период;
- построение модели осуществляется в соответствии с алгоритмом, разработанным в п. 2.3.

Источник данных: витрина централизованного хранилища данных, содержащая информацию о клиентах «Сбербанк Премьер» в разрезе клиентов, включая возраст и пол, количество и объем операций покупок по категориям, количество и объем поступлений и списаний денежных средств по источникам и получателям, наличие кредитных продуктов, остатки по депозитам и дебетовым картам, объем денежных средств, размещенных в инвестиционных и страховых продуктах и др. Данные загружены за период январь–июль 2020 г.

3.2 Определение значимых критериев классификации для включения в дерева решений

Алгоритмы машинного обучения на основе деревьев решений, такие как `DecisionTreeRegressor` или их эквивалент для ансамблевого обучения `RandomForestRegressor`, используют набор деревьев, которые содержат узлы, полученные в результате разделения. Основная цель этих разделений – максимально возможно уменьшить количество шумов, таких как *средняя абсолютная ошибка* (mean absolute error, MAE) и *средняя квадратическая ошибка* (mean squared error, MSE) [56].

Средняя абсолютная ошибка рассчитывается как среднее абсолютных разностей между целевыми значениями и прогнозами [20, с. 34]:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (1)$$

Средняя квадратическая ошибка рассчитывается как средний квадрат разностей между целевыми значениями и прогнозами [20, с. 34]:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2)$$

Модели, основанные на данных алгоритмах, могут рассчитать, насколько важен признак, рассчитав степень уменьшения шумов за счёт этого признака.

На основании информации, хранящейся в базе данных по клиентам «Сбербанк Премьер», была создана выборка данных, содержащая следующую информацию в разрезе клиентов (таблица 6):

Таблица 6 – Перечень полей для отбора значимых признаков³⁶

Признак	Расшифровка
Client	ID клиента
Gender	Пол
Age	Возраст
AUM	Портфель под управлением
FreeMoney	Остатки на банковских картах
Deposit	Остатки на депозитах
Invest	Остаток по инвестиционным продуктам
Insur	Остаток по страховым продуктам
ISG_Part_AUM	Доля ИСЖ в портфеле клиента
ISG_Part_Insur	Доля ИСЖ в страховом портфеле
NSG_Part_AUM	Доля НСЖ в портфеле клиента
NSG_Part_Insur	Доля НСЖ в страховом портфеле
PIF_Part_AUM	Доля ПИФ в портфеле клиента
PIF_Part_Invest	Доля ПИФ в инвестиционном портфеле
DU_Part_AUM	Доля ДУ в портфеле клиента
DU_Part_Invest	Доля ДУ в инвестиционном портфеле
BO_Part_AUM	Доля БО в портфеле клиента
BO_Part_Invest	Доля БО в инвестиционном портфеле

По всем числовым признакам, кроме возраста, были рассчитаны медианные значения за последнее полугодие. Данная выборка была загружена в датафрейм Python. Фрагмент датафрейма представлен на рисунке 31.

	Client	Gender	Age	AUM	FreeMoney	Deposit	Invest	Insur	ISG_Part_AUM	ISG_Part_Insur	NSG_Part_AUM	NSG_Part_Insur
0	192218453	0	50.0	88.630	0.000	88.630	0.000	0.0	0.0000	0.0000	0.0000	0.0000
1	59353621	0	65.0	4476368.665	0.000	3289041.585	54892.895	1132650.0	0.2234	0.8829	0.0296	0.1171
2	276821586	1	37.0	115837.025	4147.255	110181.375	0.000	0.0	0.0000	0.0000	0.0000	0.0000
3	129793929	0	40.0	536064.930	117102.255	420029.360	0.000	0.0	0.0000	0.0000	0.0000	0.0000
4	209034621	0	51.0	1440428.455	663115.840	120.625	754444.235	0.0	0.0000	0.0000	0.0000	0.0000
5	119106181	1	35.0	64952.580	64952.580	0.000	0.000	0.0	0.0000	0.0000	0.0000	0.0000

Рисунок 31 – Фрагмент исходного датафрейма для отбора признаков³⁷

Рассмотрим методику отбора значимых признаков на примере ИСЖ:

1. В качестве целевой переменной рассматривается доля ИСЖ в суммарном портфеле клиента. Данное поле исходного датафрейма (ISG_Part_AUM) сохраняется в отдельный массив. Далее из исходного датафрейма удаляется поле Client, поскольку это просто код клиента в системе, а также удаляются целевые переменные для других продуктов (NSG_Part_AUM, PIF_Part_AUM, DU_Part_AUM, BO_Part_AUM), чтобы исключить их взаимное влияние друг

³⁶ Составлено автором по: [56]

³⁷ Составлено автором по: [56]

на друга. Оставшиеся поля сохраняются в массив, который будет использован для обучения алгоритма.

2. На следующем этапе вызывается объект `RandomForestRegressor` из библиотеки `Sklearn`, которому передаются два массива: входные переменные и целевая переменная. Алгоритм обучается на этих данных.
3. В процессе обучения алгоритм рассчитывает степень влияния каждой входной переменной на целевую, выраженную в коэффициенте важности каждой входной переменной. Эти коэффициенты хранятся в атрибуте `feature_importances_` объекта `RandomForestRegressor`. Обратившись к этому атрибуту, можно получить оценку важности каждой переменной (таблица 7).

Таблица 7 – Коэффициенты значимости признаков для прогноза доли ИСЖ³⁸

Feature	Importance
ISG_Part_Insur	0.677893
Deposit	0.161856
Invest	0.071970
Insur	0.058534
FreeMoney	0.014007
AUM	0.012830
NSG_Part_Insur	0.002615
Age	0.000188
BO_Part_Invest	0.000034
Gender	0.000030
DU_Part_Invest	0.000023
PIF_Part_Invest	0.000018

Полученные коэффициенты для наглядности визуализируются средствами библиотеки `Matplotlib` (рисунок 32):

³⁸ Составлено автором по: [56]

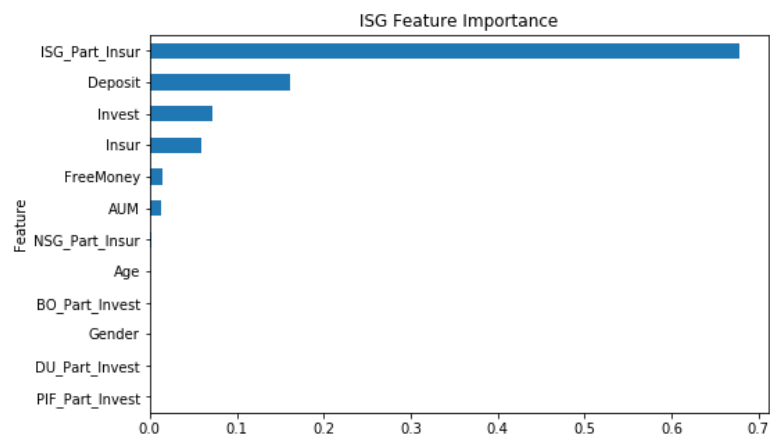


Рисунок 32 – Визуализация значимости признаков для прогноза ИСЖ³⁹

Таким образом, наиболее значимыми признаками для прогнозирования доли ИСЖ в портфеле клиента являются доля ИСЖ в страховом портфеле, остатки на депозитах, остатки по инвестиционным продуктам, остатки по страховым продуктам, остатки на банковских картах и суммарный размер портфеля под управлением. Именно на этих признаках в дальнейшем будет построена модель прогнозирования доли ИСЖ.

Аналогичная процедура проводится для каждого комиссионного продукта.

Так, в таблице 8 представлены коэффициенты значимости для доли ИСЖ:

Таблица 8 – Коэффициенты значимости признаков для прогноза доли ИСЖ⁴⁰

Feature	Importance
NSG_Part_Insur	0.810884
Deposit	0.107448
Invest	0.034805
Insur	0.032375
FreeMoney	0.007736
AUM	0.004536
ISG_Part_Insur	0.001904
Age	0.000198
BO_Part_Invest	0.000041
Gender	0.000031
DU_Part_Invest	0.000024
PIF_Part_Invest	0.000017

Значимость признаков визуализирована на рисунке 33:

³⁹ Составлено автором по: [56]

⁴⁰ Составлено автором по: [56]

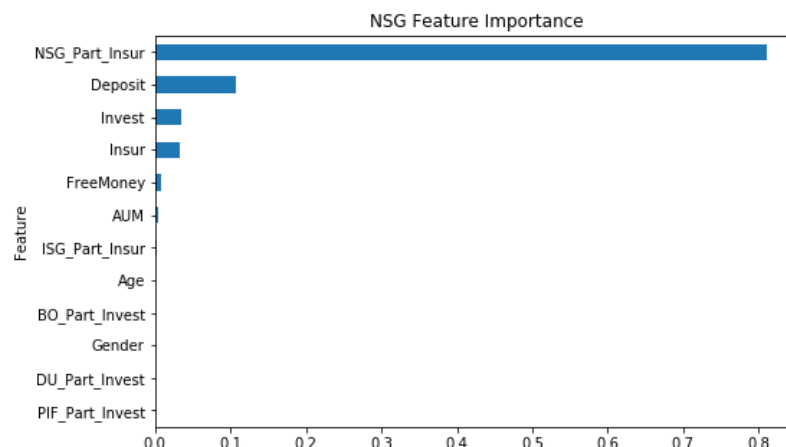


Рисунок 33 – Визуализация значимости признаков для прогноза НСЖ⁴¹

Наиболее значимыми признаками для прогнозирования доли НСЖ в портфеле клиента являются доля НСЖ в страховом портфеле, остатки на депозитах, остатки по инвестиционным продуктам, остатки по страховым продуктам, остатки на банковских картах и суммарный размер портфеля под управлением. Этих признаки будут лежать в основе модели прогнозирования доли НСЖ.

В таблице 9 представлены коэффициенты значимости для доли ПИФ:

Таблица 9 – Коэффициенты значимости признаков для прогноза доли ПИФ⁴²

Feature	Importance
PIF_Part_Invest	0.706591
Invest	0.111176
Deposit	0.088058
Insur	0.061580
AUM	0.017834
FreeMoney	0.010723
BO_Part_Invest	0.001998
Age	0.001162
ISG_Part_Insur	0.000374
Gender	0.000228
DU_Part_Invest	0.000150
NSG_Part_Insur	0.000127

Значимость признаков визуализирована на рисунке 34:

⁴¹ Составлено автором по: [56]

⁴² Составлено автором по: [56]

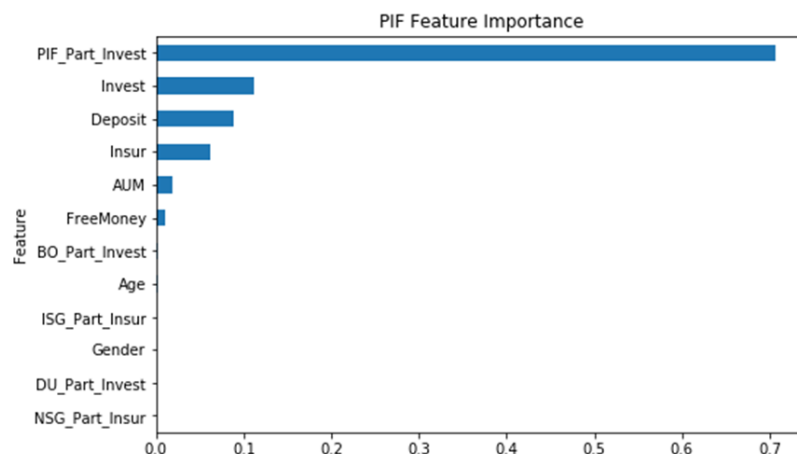


Рисунок 34 – Визуализация значимости признаков для прогноза ПИФ⁴³

Наиболее значимыми признаками для прогнозирования доли ПИФ в портфеле клиента являются доля ПИФ в инвестиционном портфеле, остатки по инвестиционным продуктам, остатки на депозитах, остатки по страховым продуктам, суммарный размер портфеля под управлением, а также остатки на банковских картах. Этих признаки будут лежать в основе модели прогнозирования доли ПИФ.

В таблице 10 представлены коэффициенты значимости для доли ДУ:

Таблица 10 – Коэффициенты значимости признаков для прогноза доли ДУ⁴⁴

Feature	Importance
DU_Part_Invest	0.687877
Deposit	0.139220
Insur	0.082387
Invest	0.061996
AUM	0.016099
FreeMoney	0.011760
BO_Part_Invest	0.000268
Age	0.000257
Gender	0.000047
ISG_Part_Insur	0.000043
NSG_Part_Insur	0.000029
PIF_Part_Invest	0.000019

Значимость признаков визуализирована на рисунке 35:

⁴³ Составлено автором по: [56]

⁴⁴ Составлено автором по: [56]

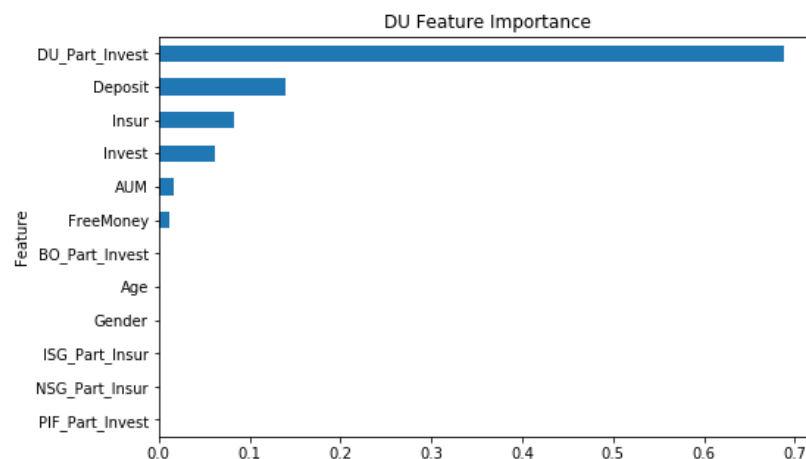


Рисунок 35 – Визуализация значимости признаков для прогноза ДУ⁴⁵

Наиболее значимыми признаками для прогнозирования доли ДУ в портфеле клиента являются доля ДУ в инвестиционном портфеле, остатки на депозитах, остатки по страховым продуктам, остатки по инвестиционным продуктам, суммарный размер портфеля под управлением, а также остатки на банковских картах.

В таблице 11 представлены коэффициенты значимости для доли БО:

Таблица 11 – Коэффициенты значимости признаков для прогноза доли БО⁴⁶

Feature	Importance
BO_Part_Invest	0.611744
Invest	0.161421
Deposit	0.107489
Insur	0.053849
AUM	0.046395
FreeMoney	0.018255
Age	0.000554
PIF_Part_Invest	0.000091
Gender	0.000073
DU_Part_Invest	0.000066
ISG_Part_Insur	0.000044
NSG_Part_Insur	0.000020

Значимость признаков визуализирована на рисунке 36:

⁴⁵ Составлено автором по: [56]

⁴⁶ Составлено автором по: [56]

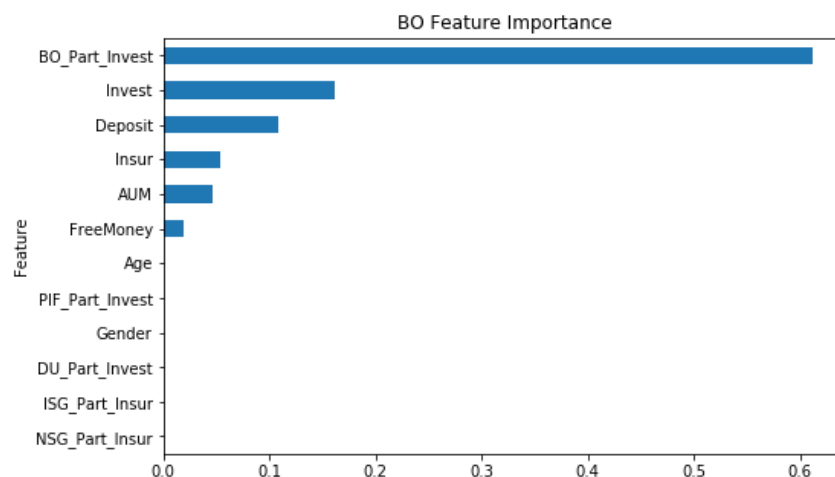


Рисунок 36 – Визуализация значимости признаков для прогноза БО⁴⁷

Наиболее значимыми признаками для прогнозирования доли БО в портфеле клиента являются доля БО в инвестиционном портфеле, остатки по инвестиционным продуктам, остатки на депозитах, остатки по страховым продуктам, суммарный размер портфеля под управлением, а также остатки на банковских картах.

3.3 Этапы обработки данных

1. Подготовка обучающей выборки.

В исходную выборку данных для обучения модели включаются все признаки, имеющие значимость для прогнозирования той или иной целевой переменной, а также сами целевые переменные (таблица 12). По каждому полю рассчитывается медианное значение за период январь–июнь 2020 г.

⁴⁷ Составлено автором по: [56]

Таблица 12 – Перечень полей для формирования обучающей выборки⁴⁸

Признак	Расшифровка
AUM	Портфель под управлением
FreeMoney	Остатки на банковских картах
Deposit	Остатки на депозитах
Invest	Остаток по инвестиционным продуктам
Insur	Остаток по страховым продуктам
ISG_Part_AUM	Доля ИСЖ в портфеле клиента
ISG_Part_Insur	Доля ИСЖ в страховом портфеле
NSG_Part_AUM	Доля НСЖ в портфеле клиента
NSG_Part_Insur	Доля НСЖ в страховом портфеле
PIF_Part_AUM	Доля ПИФ в портфеле клиента
PIF_Part_Invest	Доля ПИФ в инвестиционном портфеле
DU_Part_AUM	Доля ДУ в портфеле клиента
DU_Part_Invest	Доля ДУ в инвестиционном портфеле
BO_Part_AUM	Доля БО в портфеле клиента
BO_Part_Invest	Доля БО в инвестиционном портфеле

2. Загрузка исходной выборки для обучения в датафрейм Python.

Фрагмент датафрейма представлен на рисунке 37. Выборка содержит 1 784 345 записей.

	Client	AUM	FreeMoney	Deposit	Invest	Insur	ISG_Part_AUM	ISG_Part_Insur	NSG_Part_AUM	NSG_Part_Insur	PIF_Part_AUM
23	130381543	1313696.760	466.020	52565.680	193299.420	1099995.2	0.0000	0.0	0.8153	1.0	0.0
35	185981743	668541.245	599.480	237291.450	432254.640	0.0	0.0000	0.0	0.0000	0.0	0.0
37	191913614	700218.330	9513.240	242372.230	54354.040	396536.0	0.0000	0.0	0.5663	1.0	0.0
57	145056925	1634432.490	3137.550	210.830	439892.110	1191192.0	0.0000	0.0	0.7288	1.0	0.0
69	431515439	5133142.805	13165.535	1103.185	3122230.895	2000000.0	0.3896	1.0	0.0000	0.0	0.0

Рисунок 37 – Фрагмент исходного датафрейма для обучения модели⁴⁹

3. Подготовка выборки для прогнозирования.

В данную выборку включаются те же поля, по каждому из которых выгружаются значения на начало текущего полугодия (на 01.07.2020), а также поля, содержащие информацию об офисе обслуживания.

4. Загрузка исходной выборки для прогнозирования в датафрейм.

⁴⁸ Составлено автором по: [58]

⁴⁹ Составлено автором по: [58]

Фрагмент датафрейма представлен на рисунке 38. Выборка содержит 1 522 243 записей.

	Client	TB	GOSB	VSP	AUM	FreeMoney	Deposit	Invest	Insur	ISG_Part_Insur	NSG_Part_Insur	PIF_Part_Invest
7	154629873	13	8596	00200	4081545.28	264025.55	2094834.67	122685.06	1600000.0	0.2500	0.7500	0.0000
92	138190440	42	8612	00500	871279.91	96113.97	465165.94	0.00	310000.0	0.3226	0.6774	0.0000
102	178646457	18	8586	00262	1404979.11	497055.51	57923.60	0.00	850000.0	0.7059	0.2941	0.0000
141	37167499	16	5940	00100	2627687.59	127681.95	3633.85	1781868.69	714503.1	0.1400	0.8600	0.1165
165	157506522	54	8621	00776	3764114.99	51590.81	23.18	0.00	3712501.0	0.7111	0.2889	0.0000

Рисунок 38 – Фрагмент исходного датафрейма для прогнозирования⁵⁰

5. Обучение модели для каждого комиссионного продукта и прогноз его доли в портфеле клиента.

Для обучения модели по ИСЖ из исходного датафрейма выбираются признаки, отобранные в п. 3.1 для прогнозирования данного продукта. Целевая переменная (ISG_Part_AUM) сохраняется в отдельный массив. Далее с помощью функции `train_test_split` массивы разделяются на обучающий (70%) и тестовый (30%), на котором в дальнейшем будет проверено качество построенной модели с помощью *коэффициента детерминации*, также известного как *R-квадрат*. Данный коэффициент всегда относителен и лежит в диапазоне от 0 до 1. Чем лучше модель умеет предсказывать данные, тем ближе значение этого коэффициента к 1 [7, с. 149].

Далее обучающая выборка передается объекту `RandomForestRegressor`, а конкретно в его метод `fit`. Глубина дерева ограничивается для исключения переобученности, и модель обучается на этих данных.

По завершении обучения вызывается метод `score`, которому передается тестовая выборка. Метод возвращает значение коэффициента детерминации. Для модели ИСЖ он равен 0.9692, что говорит о высоком качестве обучения.

Для прогнозирования доли ИСЖ в портфеле клиента из исходного датафрейма с данными на начало полугодия выбираются те же значимые признаки, и полученный массив передается в метод `predict`, который

⁵⁰ Составлено автором по: [58]

рассчитывает значения целевой переменной ISG_Part_AUM. Полученные значения сохраняются в отдельный массив.

Данный этап проходит несколько итераций, в каждой из которых строится модель для своего комиссионного продукта. Коэффициенты детерминации для всех моделей представлены в таблице 13. Все модели характеризуются высоким качеством.

Таблица 13 – Оценка качества моделей прогнозирования целевых переменных⁵¹

Модель	Коэффициент детерминации
ИСЖ	0.9692
НСЖ	0.9845
ПИФ	0.9530
ДУ	0.9709
БО	0.9643

6. Расчет плановых объемов средств клиентов, размещенных в каждом комиссионном продукте.

Прогнозные доли, рассчитанные на предыдущем этапе, умножаются на размер портфеля клиента на начало полугодия, и таким образом рассчитываются плановые объемы средств по каждому комиссионному продукту. Полученные расчетные данные сохраняются в отдельный массив.

7. Расчет комиссионного дохода по каждому продукту.

Объемы средств, полученные на предыдущем этапе, умножаются на средневзвешенную ставку комиссионного дохода за 1 полугодие 2020 г. (таблица 14). Расчет ставок производится по данным управленческой отчетности канала «Сбербанк Премьер». Полученные значения сохраняются в отдельный массив.

⁵¹ Составлено автором по: [58]

Таблица 14 – Средневзвешенные ставки по продуктам за 1 полугодие 2020 г.⁵²

Модель	Средневзвешенная ставка
ИСЖ	3,6%
НСЖ	24,7%
ПИФ	2,6%
ДУ	3,3%
БО	2,5%

8. Суммирование комиссионных доходов по каждому продукту.

Значения комиссионных доходов по каждому продукту суммируются для получения итогового плана по комиссионному доходу в разрезе клиентов. Данные сохраняются в отдельный датафрейм, к которому добавляется информация об офисе обслуживания. Фрагмент итогового датафрейма представлен на рисунке 39.

	TB	GOSB	VSP	total_com
0	70	9070	00017	0.018680
1	40	8606	00001	251.507720
2	38	9038	01690	0.005930
3	42	8611	00264	0.912578
4	44	8047	00299	76.470572

Рисунок 39 – Фрагмент итогового датафрейма с планами по клиентам⁵³

9. Полученные планы суммируются с группировкой по территориальному банку, головному отделению и офису и выгружаются в Excel для финального форматирования и последующей отправки на согласование.

3.4 Результаты работы модели и экономическое обоснование

Для оценки результатов работы модели произведен расчет фактического выполнения плана по комиссионному доходу «Сбербанк Премьер» за период июль–декабрь 2020 года. Выполнение оценено по официальному Бизнес-плану,

⁵² Составлено автором по: [58]

⁵³ Составлено автором по: [58].

рассчитанному по текущей методике, и по планам, рассчитанным с помощью модели в п. 3.2. Результаты расчетов представлены в таблице 15 в разрезе ТБ.

Таблица 15 – Оценка выполнения планов по Бизнес-плану и по модели⁵⁴

Территориальный банк	Факт за 2 пг 2020, тыс. руб.	Бизнес-план на 2 пг 2020, тыс. руб.	Выполнение по Бизнес-плану, %	План по модели на 2 пг 2020, тыс. руб.	Выполнение по модели, %
Байкальский банк	157 378	144 504	108,9%	151 842	105,2%
Волго-Вятский банк	601 827	500 813	120,2%	533 032	114,6%
Дальневосточный банк	187 164	196 744	95,1%	186 429	101,9%
Московский банк	2 968 243	2 033 267	146,0%	2 319 297	129,9%
Поволжский банк	609 078	403 245	151,0%	454 904	135,9%
Северо-Западный банк	663 369	580 953	114,2%	630 449	106,8%
Сибирский банк	395 372	389 995	101,4%	379 662	105,7%
Среднерусский банк	911 839	732 091	124,6%	802 007	115,4%
Уральский банк	462 835	484 787	95,5%	449 118	104,6%
Центрально-Черноземный б	258 273	211 049	122,4%	221 783	118,2%
Юго-Западный банк	347 403	318 212	109,2%	328 624	107,3%
Сбербанк	7 562 781	5 995 659	126,1%	6 457 148	117,1%

Данные таблицы 15 показывают, что по сравнению с Бизнес-планом модель рассчитала больший план в абсолютном выражении – 6 457 млн руб. для Сбербанка в целом по модели против 5 996 млн руб. официального плана по Бизнес-плану. Однако при этом выполнение плана всеми ТБ составляет 100% и более, тогда как по Бизнес-плану 2 ТБ – Дальневосточный и Уральский – не выполнили план по комиссионному доходу за полугодие. Это означает, что несмотря на увеличение абсолютного значения плана, он распределяется по всем ТБ равномернее, позволяя обеспечить выполнение плана всеми ТБ и сокращая разрыв между ними.

Для оценки соответствия модели ключевому бизнес-требованию рассчитана доля офисов, выполнивших план за полугодие по Бизнес-плану и по модели. Результаты расчетов представлены в таблице 16.

⁵⁴ Составлено автором по: [58]

Таблица 16 – Доля офисов, выполнивших план по Бизнес-плану и по модели⁵⁵

Территориальный банк	Доля офисов, выполнивших Бизнес-план, %	Доля офисов, выполнивших план по модели, %
Байкальский банк	94,1%	96,3%
Волго-Вятский банк	100,0%	100,0%
Дальневосточный банк	95,8%	97,0%
Московский банк	43,8%	62,1%
Поволжский банк	100,0%	100,0%
Северо-Западный банк	95,3%	97,4%
Сибирский банк	96,0%	100,0%
Среднерусский банк	95,7%	98,3%
Уральский банк	98,6%	100,0%
Центрально-Черноземный банк	80,0%	90,2%
Юго-Западный банк	97,2%	100,0%
Сбербанк	86,6%	91,1%

Таким образом, при текущем подходе к планированию комиссионного дохода «Сбербанк Премьер» 86,6% офисов в целом по системе выполняют план на полугодие. При этом стопроцентное выполнение плана всеми офисами обеспечивается только в двух ТБ. Расчет планов с помощью модели позволяет увеличить количество ТБ, в которых план выполняется всеми офисами, до пяти, а также повысить долю офисов, выполняющих план в целом по Сбербанку, до 91,1%. Это полностью соответствует ключевому бизнес-требованию (более 90% офисов выполняют план на 100% и более).

Любой проект или инновация имеет стоимость реализации и эффект от реализации, который можно оценить в денежном выражении. Внедрение предлагаемого подхода позволит сократить количество трудозатрат, поскольку для расчета планов с помощью разработанной модели ежемесячно потребуется меньшее количество сотрудников, чем в настоящее время. Это позволит сэкономить рабочие ресурсы подразделения, а именно рабочее время сотрудников, и направить их на другие задачи. Такая экономия может быть конвертирована в положительный экономический эффект. Расчет экономической эффективности предлагаемого проекта произведен по следующей методике:

⁵⁵ Составлено автором по: [58]

1. Выделение затрат на реализацию бизнес-процесса, детализация и расчет ресурсов.

Произведена детализация работ по одному циклу расчета плана в часах. Расчет план производится двумя сотрудниками:

– аналитик, заработная плата составляет 50 000 руб./мес., или 312,5 руб./час.;

– эксперт, заработная плата составляет 60 000 руб./мес., или 375 руб./час.

Данные о времени и стоимости работ представлены в таблице 17. Затраты времени и стоимость работ до внедрения модели рассчитаны как среднее время за последнее полугодие, т.е. за шесть циклов планирования, затраты после внедрения определены методом экспертной оценки.

Таблица 17 – Затраты для каждой работы бизнес-процесса «Расчет плана по комиссионному доходу "Сбербанк Премьер"»⁵⁶

Бизнес-процесс «Расчет плана по комиссионному доходу "Сбербанк Премьер"»						
	До внедрения модели			После внедрения модели		
	Т/шт., час.	₽/шт.	Итого, руб.	Т/шт., час.	₽/шт.	Итого, руб.
Подготовка данных для расчета	72	313	22 500	40	313	12 500
Расчет планов	64	375	24 000	48	375	18 000
ИТОГО	136		46 500	88		30 500

2. Расчет и оценка затрат на каждом этапе жизненного цикла.

Для предлагаемого решения была выбрана каскадная модель жизненного цикла, включающая в себя следующие этапы: определение требований, проектирование, разработка, тестирование, эксплуатация и поддержка.

Для каждого этапа жизненного цикла на уровне надсистемы, системы и подсистемы определяются необходимые работы и исполнители, а также затраты на их выполнение во временном и стоимостном выражении (таблица 18).

⁵⁶ Составлено автором по: [5]

Таблица 18 – Работы на этапах жизненного цикла в разрезе надсистемы, системы и подсистемы⁵⁷

		Определение требований	Проектирование	Разработка	Тестирование	Эксплуатация и поддержка
		01.06.2021-06.06.2021	07.06.2021-13.06.2021	14.06.2021-22.06.2021	23.06.2021-28.06.2021	29.06.2021-30.06.2021
Надсистема	Управление планирования	Выделение бюджета, проведение установочной встречи	Не изменяется	Не изменяется	Оценка качества разработанной модели	Проверка знаний у менеджера поддержки
	Затраты, час	2			1	2
Система	Модель планирования комиссионного дохода	Разработка и согласование технического задания	Разработка комплекта проектной документации	Построение архитектуры модели	Проведение тестового расчета Проверка соответствия ТЗ	Ввод в эксплуатацию, поддержание работоспособности модели
	Затраты, час	3	5	10	5	1
Подсистемы	Модуль подготовки данных	Разработка и согласование технического задания	Анализ источников данных, выбор подходящих признаков	Разработка SQL-скриптов для сбора исходных данных	Отладка кода	Не изменяется
	Затраты, час	3	6	8	3	
	Модуль обучения	Разработка и согласование технического задания	Подбор оптимального алгоритма машинного обучения	Отбор признаков, обучение моделей в Python	Отладка кода	Не изменяется
	Затраты, час	3	4	12	3	
	Модуль расчета планов	Разработка и согласование технического задания	Разработка алгоритма расчета	Обработка полученных данных в Python, проведение расчета плановых значений	Отладка кода	Не изменяется
Затраты, час	3	3	10	2		
Персонал	Директор управления планирования, эксперт, технический писатель	Технический писатель, аналитик, ведущий аналитик, эксперт	Разработчик, аналитик, ведущий аналитик, эксперт	Директор управления планирования, тестировщик, разработчик, технический писатель	Эксперт, менеджер поддержки	
Денежные затраты, руб.	5 250,00	6 181,25	15 406,25	5 431,25	1100 + 9000 руб./мес.	
Затраты времени, час.	14	18	40	14	3 + 24 час./мес.	

⁵⁷ Составлено автором по: [5]

3. Сравнение затрат до внедрения модели с ожидаемыми затратами после внедрения.

В таблице 19 представлены данные о затратах на расчет плана по комиссионному доходу «Сбербанк Премьер» до и после внедрения модели. Ставка исполнителя рассчитана как усредненная заработная плата в час аналитика и эксперта. Стоимость поддержки модели после внедрения складывается из 24 часов в месяц работы менеджера поддержки.

Таблица 19 – Условия до и после внедрения модели планирования комиссионного дохода «Сбербанк Премьер»⁵⁸

Показатель	До внедрения модели	После внедрения модели
Расчет плана по комиссионному доходу, час.	136	88
Стоимость поддержания, руб./месяц		9 000
Ставка исполнителя, руб./час	343,75	343,75

На основе данных, представленных в таблицах 19 и 17, строится график затрат (рисунок 40). По вертикальной оси откладываются затраты в рублях, по горизонтальной – количество циклов расчета плана. На диаграмме представлены следующие данные:

- полные затраты на расчет плана по комиссионному доходу до внедрения модели планирования;
- полные затраты на расчет плана по комиссионному доходу после внедрения модели планирования;
- стоимость поддержки модели планирования комиссионного дохода после автоматизации.

По данным диаграммы можно сделать вывод, что в первом же цикле расчета плана по комиссионному доходу с помощью модели планирования ее использование становится выгодным.

⁵⁸ Составлено автором по: [5]

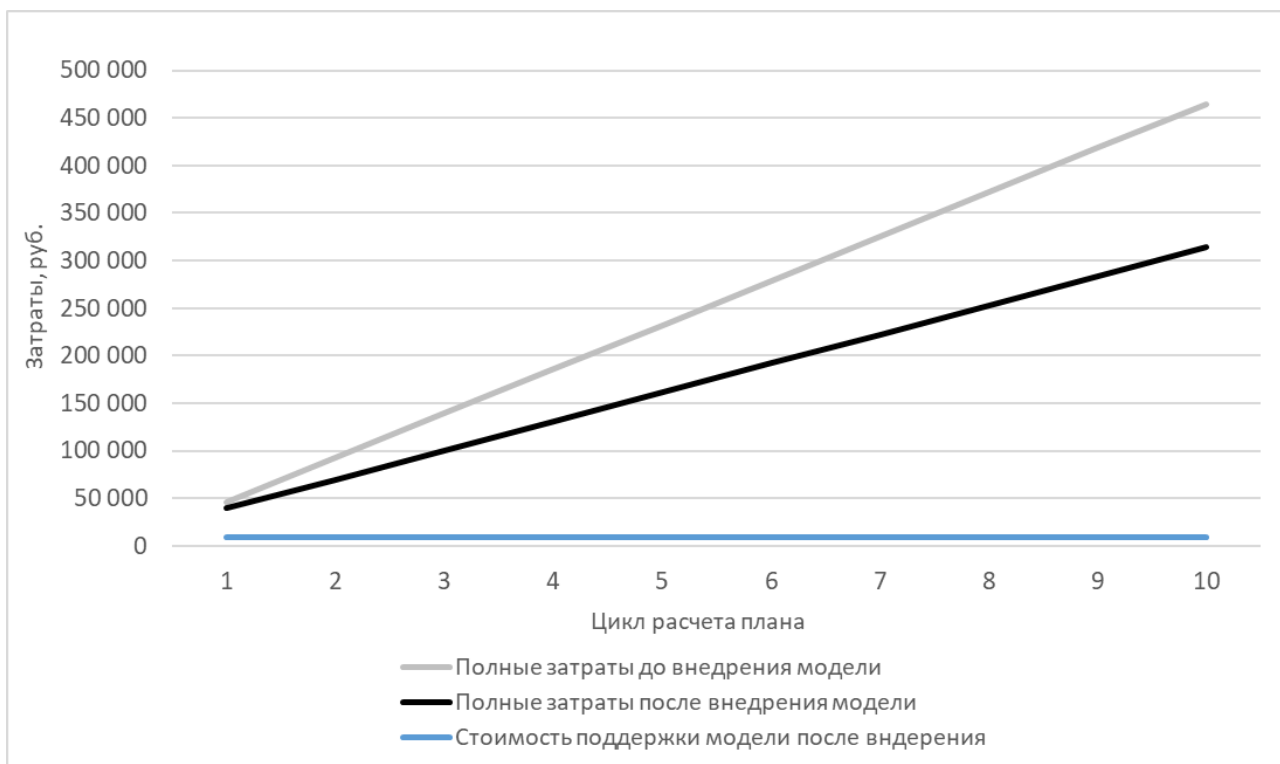


Рисунок 40 – Затраты на бизнес-процесс «Расчет плана по комиссионному доходу «Сбербанк Премьер»»⁵⁹

4. Расчет точки безубыточности

Проект внедрения включает в себя пять основных этапов (определение требований, проектирование, разработка, тестирование, эксплуатация и поддержка), каждый из которых декомпозируется на задачи:

Сетевой график

1. Начало проекта (2).
2. Определение требований:
 - выделение бюджета (4);
 - проведение установочной встречи (5);
 - разработка и согласование технического задания (6).
3. Проектирование:
 - разработка комплекта проектной документации (8);
 - анализ источников данных (9);

⁵⁹ Составлено автором по: [5]

- выбор подходящих признаков (10);
- подбор оптимального алгоритма машинного обучения (11);
- разработка алгоритма расчета (12).

4. Разработка:

- построение архитектуры модели (14);
- разработка SQL-скриптов для сбора данных (15);
- отбор признаков (16);
- обучение моделей в Python (17);
- обработка полученных данных в Python (18);
- проведение расчета плановых значений (19).

5. Тестирование:

- проведение тестового расчета (21);
- оценка качества разработанной модели (22);
- проверка соответствия ТЗ (23);
- отладка кода (24).

6. Эксплуатация и поддержка:

- проверка знаний у менеджера поддержки (26);
- ввод в эксплуатацию (27).

7. Окончание проекта (28).

На рисунке 41 представлен сетевой график. Почти все задачи проекта являются критическими. Проект выполняется в короткие сроки, поэтому очень важно строго отслеживать выполнение всех работ и обеспечивать наличие ресурсов на всех этапах проекта.

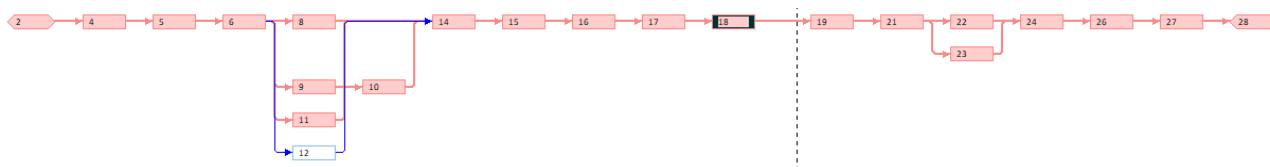


Рисунок 41 – Сетевой график проекта в Microsoft Project⁶⁰

⁶⁰ Составлено автором по: [49]

Календарное планирование проекта

Календарное планирование проводится средствами MS Project. На рисунке 42 представлен перечень работ с датой начала и окончания. Проект планируется выполнить в срок с 01.06.2021 до 30.06.2021. Общая продолжительность проекта составляет 89 часов. Наиболее затратным по времени является этап «Разработка», на выполнение которого необходимо 40 часов.

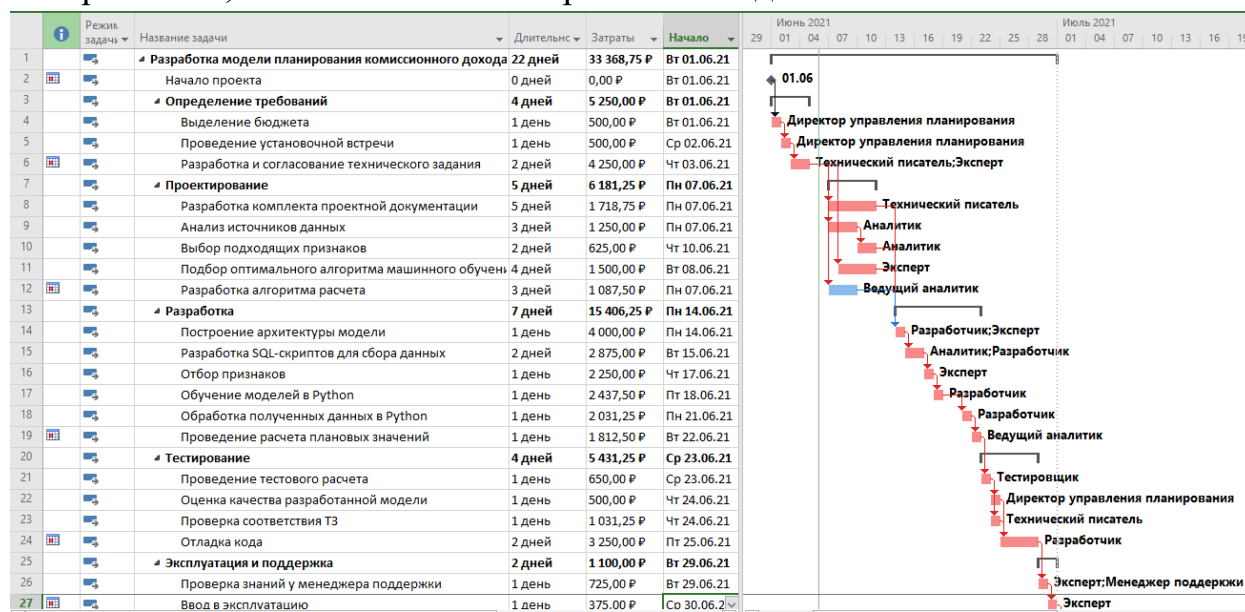


Рисунок 42 – Декомпозиция этапов работа и диаграмма Ганта⁶¹

Ресурсное планирование проекта

Проект по внедрению модели планирования выполняется командой из восьми человек:

- директор управления планирования – один человек, ставка – 500 руб./час;
- аналитик – один человек, ставка – 312,5 руб./час;
- ведущий аналитик – один человек, ставка – 362,5 руб./час;
- эксперт – один человек, ставка – 375 руб./час;
- технический писатель – один человек, ставка – 343,75 руб./час;
- разработчик – один человек, ставка – 406,25 руб./час;
- тестировщик – один человек, ставка – 325 руб./час;

⁶¹ Составлено автором по: [49]

– менеджер поддержки – один человек, ставка – 350 руб./час.

Лист ресурсов из MS Project представлен на рисунке 43.

	Название ресурса	Тип	Единицы измерения материалс	Краткое название	Группа	Макс. единиц	Стандартная ставка	Ставка сверхурочн	Затраты на исполыз.	Начисление	Базовый календарь
1	Директор управления планирования	Трудовой		Д		100%	500,00 Р/ч	1 000,00 Р/ч	0,00 Р	Пропорциональ	Стандартный
2	Аналитик	Трудовой		А		100%	312,50 Р/ч	625,00 Р/ч	0,00 Р	Пропорциональ	Стандартный
3	Ведущий аналитик	Трудовой		В		100%	362,50 Р/ч	725,00 Р/ч	0,00 Р	Пропорциональ	Стандартный
4	Эксперт	Трудовой		Э		100%	375,00 Р/ч	750,00 Р/ч	0,00 Р	Пропорциональ	Стандартный
5	Технический писатель	Трудовой		Т		100%	343,75 Р/ч	687,50 Р/ч	0,00 Р	Пропорциональ	Стандартный
6	Разработчик	Трудовой		Р		100%	406,25 Р/ч	812,50 Р/ч	0,00 Р	Пропорциональ	Стандартный
7	Тестировщик	Трудовой		Т		100%	325,00 Р/ч	650,00 Р/ч	0,00 Р	Пропорциональ	Стандартный
8	Менеджер поддержки	Трудовой		М		100%	350,00 Р/ч	700,00 Р/ч	0,00 Р	Пропорциональ	Стандартный

Рисунок 43 – Лист ресурсов проекта⁶²

На рисунке 44 представлены данные о количестве часов, затрачиваемых работником на каждую задачу, а также общее количество часов каждого исполнителя.

	Название ресурса	Трудозатрат	Подробность	Июн '21				Июл '21	
				31	07	14	21	28	05
1	Директор управления планирования	3 ч	Трудозатр.	2ч			1ч		
	Выделение бюджета	1 ч	Трудозатр.	1ч					
	Проведение установочной встречи	1 ч	Трудозатр.	1ч					
	Оценка качества разработанной модели	1 ч	Трудозатр.				1ч		
2	Аналитик	10 ч	Трудозатр.		6ч	4ч			
	Анализ источников данных	4 ч	Трудозатр.		4ч				
	Выбор подходящих признаков	2 ч	Трудозатр.		2ч				
	Разработка SQL-скриптов для сбора данных	4 ч	Трудозатр.			4ч			
3	Ведущий аналитик	8 ч	Трудозатр.		3ч		5ч		
	Разработка алгоритма расчета	3 ч	Трудозатр.		3ч				
	Проведение расчета плановых значений	5 ч	Трудозатр.				5ч		
4	Эксперт	18 ч	Трудозатр.	4ч	4ч	8ч		2ч	
	Разработка и согласование технического задания	4 ч	Трудозатр.	4ч					
	Подбор оптимального алгоритма машинного обучения	4 ч	Трудозатр.		4ч				
	Построение архитектуры модели	2 ч	Трудозатр.			2ч			
	Отбор признаков	6 ч	Трудозатр.			6ч			
	Проверка знаний у менеджера поддержки	1 ч	Трудозатр.					1ч	
	Ввод в эксплуатацию	1 ч	Трудозатр.					1ч	
5	Технический писатель	16 ч	Трудозатр.	8ч	5ч		3ч		
	Разработка и согласование технического задания	8 ч	Трудозатр.	8ч					
	Разработка комплекта проектной документации	5 ч	Трудозатр.		5ч				
	Проверка соответствия ТЗ	3 ч	Трудозатр.				3ч		
6	Разработчик	31 ч	Трудозатр.			18ч	9ч	4ч	
	Построение архитектуры модели	8 ч	Трудозатр.			8ч			
	Разработка SQL-скриптов для сбора данных	4 ч	Трудозатр.			4ч			
	Обучение моделей в Python	6 ч	Трудозатр.			6ч			
	Обработка полученных данных в Python	5 ч	Трудозатр.				5ч		
	Отладка кода	8 ч	Трудозатр.				4ч	4ч	
7	Тестировщик	2 ч	Трудозатр.				2ч		
	Проведение тестового расчета	2 ч	Трудозатр.				2ч		
8	Менеджер поддержки	1 ч	Трудозатр.					1ч	
	Проверка знаний у менеджера поддержки	1 ч	Трудозатр.					1ч	

Рисунок 44 – Лист использования ресурсов⁶³

⁶² Составлено автором по: [49]

⁶³ Составлено автором по: [49]

Стоимостное планирование проекта

На основании представленных выше данных рассчитаны плановые затраты на реализацию проекта, которые составляют 33 368,75 руб. с учетом длительности проекта 89 часов (с 01.06.2021 по 30.06.2021). На рисунке 45 представлена таблица затрат, данные которой показывают, что наиболее затратным этапом является этап «Разработка» (46% затрат).

	Название задачи	Начисление фикс. затрат	Общие затраты	Трудозатр
1	Разработка модели планирования комиссионного дохода	Пропорциональное	33 368,75 Р	89 ч
2	Начало проекта	Пропорциональное	0,00 Р	0 ч
3	Определение требований	Пропорциональное	5 250,00 Р	14 ч
4	Выделение бюджета	Пропорциональное	500,00 Р	1 ч
5	Проведение установочной встречи	Пропорциональное	500,00 Р	1 ч
6	Разработка и согласование технического задания	Пропорциональное	4 250,00 Р	12 ч
7	Проектирование	Пропорциональное	6 181,25 Р	18 ч
8	Разработка комплекта проектной документации	Пропорциональное	1 718,75 Р	5 ч
9	Анализ источников данных	Пропорциональное	1 250,00 Р	4 ч
10	Выбор подходящих признаков	Пропорциональное	625,00 Р	2 ч
11	Подбор оптимального алгоритма машинного обучения	Пропорциональное	1 500,00 Р	4 ч
12	Разработка алгоритма расчета	Пропорциональное	1 087,50 Р	3 ч
13	Разработка	Пропорциональное	15 406,25 Р	40 ч
14	Построение архитектуры модели	Пропорциональное	4 000,00 Р	10 ч
15	Разработка SQL-скриптов для сбора данных	Пропорциональное	2 875,00 Р	8 ч
16	Отбор признаков	Пропорциональное	2 250,00 Р	6 ч
17	Обучение моделей в Python	Пропорциональное	2 437,50 Р	6 ч
18	Обработка полученных данных в Python	Пропорциональное	2 031,25 Р	5 ч
19	Проведение расчета плановых значений	Пропорциональное	1 812,50 Р	5 ч
20	Тестирование	Пропорциональное	5 431,25 Р	14 ч
21	Проведение тестового расчета	Пропорциональное	650,00 Р	2 ч
22	Оценка качества разработанной модели	Пропорциональное	500,00 Р	1 ч
23	Проверка соответствия ТЗ	Пропорциональное	1 031,25 Р	3 ч
24	Отладка кода	Пропорциональное	3 250,00 Р	8 ч
25	Эксплуатация и поддержка	Пропорциональное	1 100,00 Р	3 ч
26	Проверка знаний у менеджера поддержки	Пропорциональное	725,00 Р	2 ч
27	Ввод в эксплуатацию	Пропорциональное	375,00 Р	1 ч

Рисунок 45 – Таблица затрат⁶⁴

Согласно сформированному в MS Project фрагменту отчета о движении денежных средств (рисунок 46), затраты на проект внедрения составят в июне 2021 года 33 368,75 руб.

⁶⁴ Составлено автором по: [49]



Рисунок 46 – Фрагмент отчета о движении денежных средств⁶⁵

Первый расчет плана по комиссионному доходу с помощью модели пройдет в июле 2021 года. В таблице 20 представлена разница затрат на расчет плана до и после внедрения модели.

Таблица 20 – Разница затрат на расчет плана по комиссионному доходу «Сбербанк Премьер» до и после внедрения модели⁶⁶

Месяц	Количество циклов расчета плана	Затраты до внедрения модели, руб.	Затраты после внедрения модели, руб.	Разница затрат, руб.
ИЮЛ 2021	1	46 500	39 500	7 000
АВГ 2021	1	46 500	39 500	7 000
СЕН 2021	1	46 500	39 500	7 000
ОКТ 2021	1	46 500	39 500	7 000
НОЯ 2021	1	46 500	39 500	7 000
ДЕК 2021	1	46 500	39 500	7 000
ЯНВ 2022	1	46 500	39 500	7 000
ФЕВ 2022	1	46 500	39 500	7 000
МАР 2022	1	46 500	39 500	7 000
АПР 2022	1	46 500	39 500	7 000
МАЙ 2022	1	46 500	39 500	7 000
ИЮН 2022	1	46 500	39 500	7 000

⁶⁵ Составлено автором по: [49]

⁶⁶ Составлено автором по: [5]

По данным таблицы 20 и рисунка 46 были построены:

- график «Затраты и экономия в месяц» (рисунок 47), который показывает, что максимум затрат приходится на июнь 2021 года и составляет 33 368,75 руб.;
- график «Затраты и экономия с накоплением» (рисунок 48), который иллюстрирует экономический эффект от внедрения модели планирования.

Начиная с ноября 2021 года экономия от внедрения модели планирования комиссионного дохода станет прибылью для компании, а ожидаемая прибыль от внедрения на 01.07.2022 составит 50 631,25 руб.

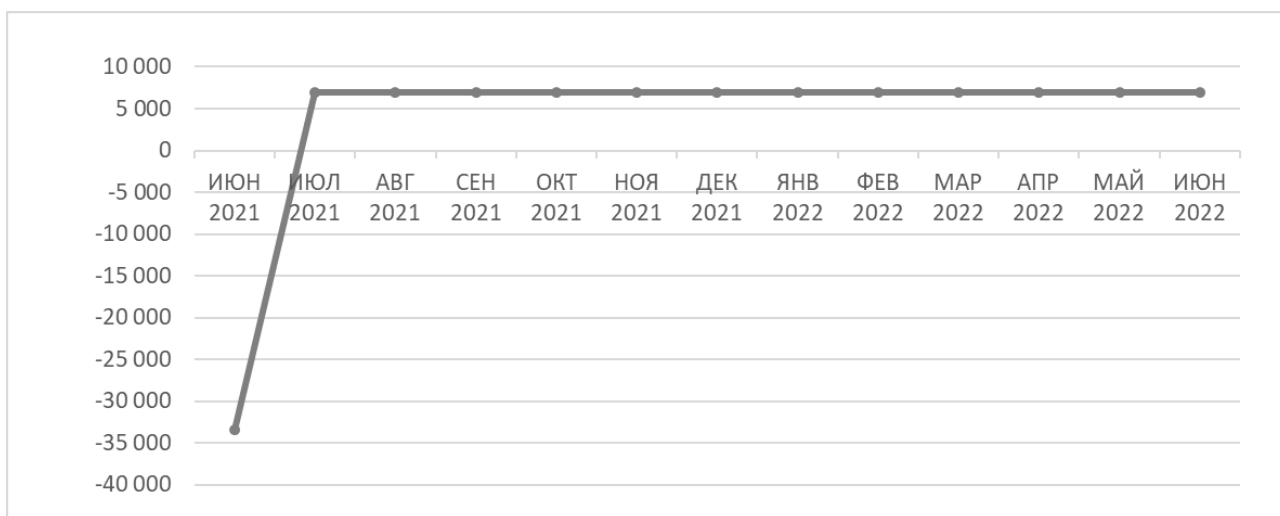


Рисунок 47 – Затраты и экономия в месяц⁶⁷

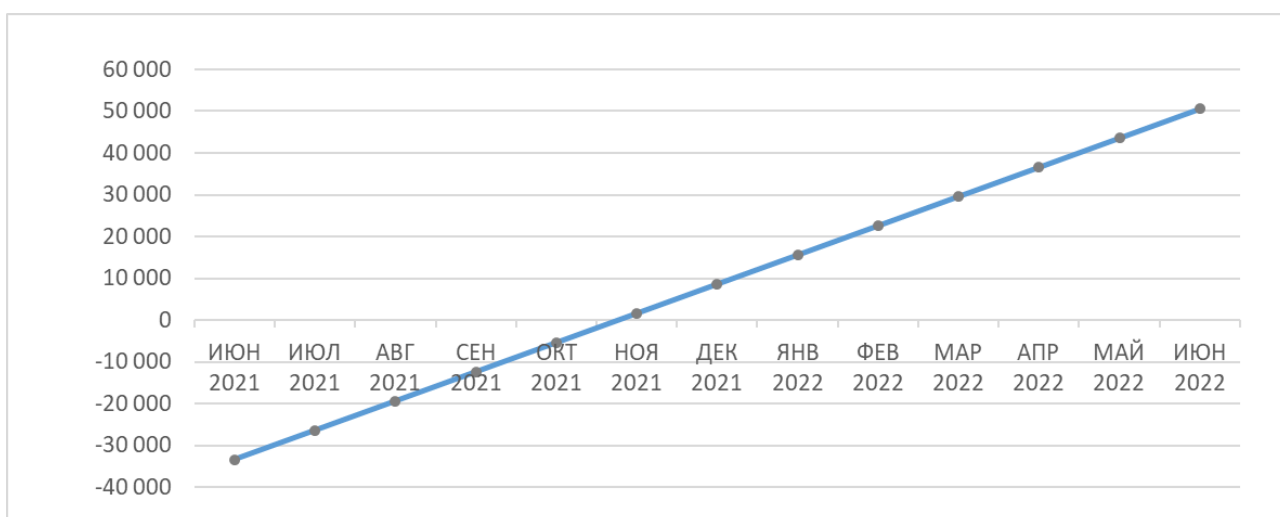


Рисунок 48 – Затраты и экономия с накоплением⁶⁸

⁶⁷ Составлено автором по: [5]

⁶⁸ Составлено автором по: [5]

5. Суммирование результатов проведенной оценки экономической эффективности проекта

Управление планирования проводит внедрение модели планирования комиссионного дохода «Сбербанк Премьер», основанной на применении методов машинного обучения. Проект имеет следующие ограничения:

- бюджет проекта составляет 50 000 руб. на все производство;
- команда проекта состоит из восьми человек;
- все работы по внедрению модели должны быть завершены в течение 1 месяца.

Внедрение модели позволит сократить расходы времени на расчет планов на 48 часов, или на 16 000 руб. в денежном выражении.

Помимо работ на различных этапах жизненного цикла модели, необходимо также провести работы на уровнях модуля подготовки данных, модуля обучения и модуля расчета планов.

Стоимость поддержки модели после внедрения составляет 24 часа в месяц работы менеджера поддержки, или 9 000 руб. в денежном выражении. Использование модели планирования будет выгодно уже в первом цикле расчета планов после ее внедрения.

Календарное планирование проекта выполнено в MS Project, выполнение проекта запланировано с 01.06.2021 по 30.06.2021, общая продолжительность – 89 часов. В проекте участвуют директор управления планирования, аналитик, ведущий аналитик, эксперт, технический писатель, разработчик, тестировщик и менеджер поддержки.

Общие затраты на внедрение модели планирования составят 33 368,75 руб.

Проведенные расчеты показывают, что внедрение модели планирования комиссионного дохода позволит экономить в среднем 7 000 руб./мес. Начиная с ноября 2021 года экономия от внедрения модели планирования превысит затраты на реализацию проекта по ее внедрению. Ожидаемая прибыль от внедрения модели составит 50 631,25 руб. на 01.07.2022.

Таким образом, модель планирования комиссионного дохода, разработанная с использованием методов машинного обучения, дает более точные и качественные результаты прогнозирования, требует меньших затрат человеческих ресурсов и времени, чем стандартные статистические методы прогнозирования, а также является экономически выгодным проектом. Автор считает необходимым отметить и такое преимущество модели, как более простое по сравнению с текущим процессом проведение корректировки плановых значений с учетом оперативных данных о клиентских операциях, что позволяет более гибко реагировать на изменение рыночной ситуации.

В числе основных рекомендаций по дальнейшему совершенствованию процесса планирования комиссионного дохода «Сбербанк Премьер» можно предложить разработку более сложных моделей с использованием большего количества параметров и более детализированного продуктового ассортимента. Кроме того, необходимо поддержание качества исходных данных на высоком уровне, их регулярное обновление, увеличение глубины данных, дальнейшее расширение перечня показателей. Все это позволит алгоритмам более качественно проводить обучение моделей и строить более точные предсказания, что, в конечном счете, повысит качество планов и позволит сотрудникам Банка выстраивать оптимальную стратегию продаж.

Результаты и выводы третьей главы

В третьей главе разработана модель планирования комиссионного дохода «Сбербанк Премьер» на основе регрессионных деревьев решений, описана методика выбора значимых критериев для прогнозирования целевых переменных, проведено экономическое обоснование реализации проекта и даны рекомендации по дальнейшему совершенствованию процесса планирования комиссионного дохода.

ЗАКЛЮЧЕНИЕ

Основными итогами работы, проведенной в рамках диссертационного исследования, являются:

- изучение и систематизация автором материалов современной научной литературы по тематике планирования и прогнозирования в кредитных организациях;
- анализ и структурирование информации о существующих подходах к планированию ключевых показателей эффективности деятельности кредитной организации;
- исследование современных и передовых методов машинного обучения, обзор практического опыта их применения в банковской сфере;
- построение полной модели бизнес-архитектуры ПАО Сбербанк, анализ основных ее элементов, стратегии развития, основных ценностей, миссии и целей;
- построение модели бизнес-процесса планирования комиссионного дохода AS-IS, определение основных недостатков существующего процесса и системный анализ проблемы;
- разработка методики реализации проекта по совершенствованию существующего бизнес-процесса, построение модели процесса TO-BE;
- определение шагов, необходимых для выбора из имеющегося массива данных о клиентах существенных признаков для построения модели планирования комиссионного дохода;
- практическая реализация модели планирования комиссионного дохода на основе метода случайного леса регрессионных деревьев решений, оценка качества разработанной модели;
- оценка экономической эффективности внедрения предложенной модели, разработка рекомендаций по оптимизации и дальнейшему совершенствованию процесса планирования комиссионного дохода.

Поставленная автором цель, состоящая в обосновании внедрения методов машинного обучения в бизнес-процесс планирования комиссионного дохода клиентского сегмента банка, была достигнута. Автором была предложена конкретная модель планирования комиссионного дохода клиентского сегмента «Сбербанк Премьер» с использованием современных методов машинного обучения. Эффективность предложенной модели планирования была доказана: модель обладает более высокой точностью прогнозирования и требует меньших человеческих и временных ресурсов.

В свою очередь, высокая точность прогноза ключевых показателей эффективности банка позволяет:

- лицам, принимающим управленческие решения, разрабатывать стратегический план развития перспективных направлений банковской деятельности;
- руководителям среднего звена видеть краткосрочную перспективу развития событий сроком до года, разрабатывать тактические шаги по достижению целей, проводить грамотную координацию сотрудников;
- сотрудникам подразделения планирования с меньшими затратами производить корректировку плановых значений с учетом меняющейся конъюнктуры рынка;
- сотрудникам подразделения отчетности и финансового управления точнее оценивать будущую доходность деятельности банка;
- сотрудникам подразделения мотивации качественнее проводить оценку выполненной работы и премирование сотрудников банка.

Таким образом, в ходе работы была достигнута цель и решены все поставленные задачи.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Кодекс корпоративного управления ПАО Сбербанк (утвержден решением Наблюдательного совета, протокол №42 от 08.12.2020, действующая редакция).
2. Устав ПАО Сбербанк (утвержден годовым Общим собранием акционеров, протокол №32 от 29.05.2019, действующая редакция).
3. Акулов В.Б., Рудаков М.Н. Теория организации: учебное пособие [Текст]. Петрозаводск: ПетрГУ, 2002. 122 с.
4. Белоглазова, Г.Н. Банковское дело. Организация деятельности коммерческого банка [Текст] / Г.Н. Белоглазова, Л.П. Кроливецкая. М: Издательство Юрайт, 2015. 149 с.
5. Берг Д.Б. Модели жизненного цикла: учебное пособие [Текст] / Екатеринбург: Изд-во Урал. ун-та, 2014. 74 с.
6. Боскетти А., Массарон Л., Шарден Б. Крупномасштабное машинное обучение вместе с Python [Текст]. М.: ДМК Пресс, 2018. 358 с.
7. Бринк Х., Ричардс Д., Феверолф М. Машинное обучение [Текст]. СПб.: Питер, 2017. 336 с.
8. Бурков В.Н., Новиков Д.А. Теория активных систем: состояние и перспективы [Текст]. М.: Синтег, 1999. 128 с.
9. Вайолино Б. Машинное обучение: методы и способы [Текст] // Директор информационной службы. 2018. №5. С. 42–44.
10. Васильев Е.П., Орешков В.И. Совершенствование процесса принятия управленческих решений в экономике и бизнесе на основе применения интеллектуального анализа данных [Текст] // Фундаментальные исследования. 2012. №9 (вып. 4). С. 965–971.
11. Вьюгин В.В. Математические основы теории машинного обучения и прогнозирования [Текст]. М.: МЦНМО, 2013. 387 с.
12. Голуб Н.И. Data Mining-технологии в стратегическом анализе деятельности коммерческих банков [Текст] // Реформы в России и проблемы

управления: матер. 23-й Всероссийской научной конференции молодых ученых и студентов. М.: ГУУ, 2008. №3. С. 156–163.

13. Голуб Н.И. Методология сегментирования частных клиентов банка [Текст] // Актуальные проблемы управления. 2008. №13. С. 76–80.

14. Голуб Н.И. Оценка конкурентной среды рынка банковских услуг с помощью технологии Data Mining [Текст] // Сборник научных работ. Экономика. Управление. Культура. 2007. №14. С. 85–91.

15. Голуб Н.И. Оценка потенциала доходности групп клиентов коммерческого банка на основе использования DM-технологий [Текст] // Вестник университета. 2008. №7. С. 63–68.

16. Голубева Т.Б. Основы моделирования и оптимизации процессов и систем сервиса: учебное пособие [Текст] / Т.Б. Голубева. Екатеринбург: УрФУ, 2017. 108 с.

17. Домингос П. Верховный алгоритм: как машинное обучение изменит наш мир [Текст]. М.: Манн, Иванов и Фербер, 2016. 336 с.

18. Достов В.Л. Электронные финансы. Мифы и реальность [Текст] / В.Л. Достов, П.М. Шуст, А.А. Валинурова, А.В. Пухов. М.: КНОРУС, 2012. 232 с.

19. Дубровин И.А. Бизнес-планирование на предприятии [Текст]. М.: Дашков и К, 2017. 432 с.

20. Зайдель А.Н. Элементарные оценки ошибок измерений [Текст]. Ленинград: Наука, 1967. 88 с.

21. Игнатенко А.Н. Гипертекст как инструментальное средство поддержки процесса дистанционного обучения [Текст] // Системные проблемы надежности, качества, информационных и электронных технологий: матер. междунар. конф. М.: Радио и связь, 2002. Ч. 3. С.54–55.

22. Игнатенко А.Н. Разработка концептуальной модели советующей информационной системы для анализа платежеспособности заемщиков [Текст] // Системные проблемы надежности, качества, информационных и электронных технологий: матер. междунар. конф. М.: Радио и связь, 2005. Ч. 8. С.60–64.

23. Игнатенко А.Н. Разработка структуры советующей информационной системы для анализа платежеспособности заемщиков [Текст] // Системные проблемы надежности, качества, информационных и электронных технологий: матер. междунар. конф. М.: Радио и связь, 2005. Ч. 8. С.53–59.

24. Исаев Р. А. Секреты успешных банков: бизнес-процессы и технологии [Текст]. М.: НИЦ ИНФРА-М, 2015. 222 с.

25. Исикава К. Японские методы управления качеством [Текст]. М.: Экономика, 1988. 199 с.

26. Комаров В.Ф., Волкова Л.В., Пирогов В.И. Проблемно-ориентированная методика анализа и решения организационно-экономических задач [Текст] // ЭКО. 2008. №7. С. 97–109.

27. Куницына Н. Н. Бизнес-планирование в коммерческом банке: Учебное пособие [Текст] / Н.Н. Куницына, А.В. Малеева, Л.И. Ушвицкий. М.: НИЦ ИНФРА-М, 2014. 384 с.

28. Леффингуэлл Д., Уидриг Д. Принципы работы с требованиями к программному обеспечению. Унифицированный подход [Текст]. М.: Вильямс, 2002. 448 с.

29. Мирджалили В., Рашка С. Python и машинное обучение [Текст]. СПб.: Диалектика, 2020. 848 с.

30. Миронов А.М. Машинное обучение, часть 1 [Текст]. М.: МАКС Пресс, 2018. 90 с.

31. Николаева Ю.В. Метод генерации выходной выборки для задачи нейросетевой классификации рыночных ситуаций [Текст] // Вестник Иркутского государственного технического университета. 2017. №12 (т. 21). С. 86–92.

32. Николаева Ю.В. Проектирование интеллектуальной системы поддержки принятия решений трейдеров на финансовых рынках [Текст] // Новая наука: стратегии и векторы развития, часть 1. 2017. №5. С. 75–79.

33. Орешков В.И. Интеллектуальный анализ данных как важнейший инструмент формирования интеллектуального капитала организаций [Текст] // Креативная экономика. 2011. №12. С. 84–89.

34. Орешков В.И. Интеллектуальный анализ данных как современный инструмент поддержки принятия решений в экономике и бизнесе [Текст] // European Social Science Journal. 2012. №9 (том 2). С. 482–490.

35. Писарева О.М., Голуб Н.И. Применение технологий интеллектуального анализа данных в решении задач стратегического развития коммерческого банка [Текст] // Вестник университета. 2008. №3. С. 71–75.

36. Рассел С., Норвиг П. Искусственный интеллект и универсальное мышление [Текст]. М: Вильямс, 2016. 1408 с.

37. Розенблатт Ф. Принципы нейродинамики. Перцептроны и теория механизмов мозга [Текст]. М.: Мир, 1965. 480 с.

38. Рудько-Силиванов В.В. Прогнозирование финансовых показателей деятельности кредитной организации [Текст] // Деньги и кредит, 2013. №2. С. 52–58.

39. Рындина С.В. Методы и средства моделирования бизнес-процессов: методология ARIS: учеб.-метод. пособие [Текст]. Пенза: Изд-во ПГУ, 2018. 52 с.

40. Самсонова М.В., Ефимов В.В. Технология и методы коллективного решения проблем: учебное пособие [Текст]. Ульяновск: УлГТУ, 2003. 152 с.

41. Силич В.А. Системный анализ и исследование операций: учебное пособие [Текст] / В.А. Силич, М.П. Силич. Томск: ТПУ, 2000. 97 с.

42. Спивак В.А. Управление изменениями: учебник для академического бакалавриата [Текст] / В.А. Спивак. М.: Издательство Юрайт, 2016. 357 с.

43. Тимонин А. Ю. Исследование процесса анализа текстовых и мультимедиа данных социального профиля из открытых источников информации [Текст] // Известия высших учебных заведений. Поволжский регион. Технические науки. 2017. №2. С. 19–28.

44. Тимонин А. Ю. Исследование процесса идентификации человека в сетях открытого доступа и построения его социального профиля на основе технологий Big Data [Текст] // Модели, системы, сети в экономике, технике, природе и обществе. 2016. №2. С. 112–119.

45. Тимонин А. Ю. Методы анализа гетерогенных данных для построения социального профиля [Текст] // Russian Journal of Management. 2017. №3 (vol. 5). С. 481–489.

46. Тимонин А. Ю. Теоретико-множественное математическое моделирование состояния социальной среды [Текст] // Известия высших учебных заведений. Поволжский регион. Технические науки. 2019. №1. С.18–33.

47. Трофимова Л.А. Методы принятия управленческих решений: учебное пособие [Текст] / Л.А. Трофимова, В.В. Трофимов. СПб.: Изд-во СПбГУЭФ, 2012. 101 с.

48. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных [Текст]. М.: ДМК Пресс, 2015. 400 с.

49. Экономическая эффективность технических решений: учебное пособие [Текст] / под ред. И.В. Ершовой. Екатеринбург: Изд-во Урал. ун-та, 2016. 140 с.

50. Экономический анализ деятельности коммерческого банка: Учебное пособие [Текст] / Ю.Г. Вешкин, Г.Л. Авагян. М.: НИЦ ИНФРА-М, 2014. 432 с.

51. Agrawal R., Imielinski T., Swami A. Mining association rules between sets of items in large databases [Текст] // Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, 1993. P. 207–216.

52. Friedman J., Hastie T., Tibshirani R. The Elements of Statistical Learning. Data Mining, Inference and Prediction [Текст]. NY.: Springer, 2009. 745 с.

53. Kohonen T. Self-Organizing Maps [Текст]. Berlin – New York: Springer-Verlag, 1989. 502 p.

54. Quinlan J.R. Induction of Decision Trees [Текст] // Machine Learning. 1986. №1. P. 81–106.

55. Samuel A.L. Some studies in machine learning using the game of checkers [Текст] // IBM Journal of Research and Development. 1959. №3. P. 210–229.

56. Костерин В.В. Отбор признаков с помощью Scikit-Learn в Python [Электронный ресурс]. URL: <https://waksoft.susu.ru/2021/02/09/otbor-priznakov-s-pomoshhyu-scikit-learn-v-python> (дата обращения: 01.05.2021).

57. Николаева Ю.В. Методы нейросетевой классификации ситуаций на финансовых рынках относительно ожидаемой прибыли [Электронный ресурс] // материалы IV Всероссийской научно-технической конференции аспирантов, магистрантов и молодых ученых с международным участием. Ижевск: ИННОВА, 2016. С. 562–566. 1 электрон. опт. диск (CD-ROM).

58. Открытый курс машинного обучения [Электронный ресурс] URL: <https://habr.com/ru/company/ods/blog/322626/> (дата обращения: 01.05.2021).

59. Официальный сайт ПАО Сбербанк [Сайт]. URL: <https://www.sberbank.com> (дата обращения: 26.04.2021).

60. ПАО Сбербанк. Стратегия 2023 [Презентация]. URL: https://sberbank.com/common/img/uploaded/files/info/sber_investor_day-strategy_2023_ru.pdf (дата обращения: 26.04.2021).

61. Тебайкина Н.И. Бизнес-модель компании: учеб.-метод. пособие [Презентация]. УрФУ, 2015. Слайды 10–35.

62. Violino B. Machine Learning: When to Use Each Method and Technique [Электронный ресурс] // InfoWorld, 2018. URL: <https://www.infoworld.com/article/3300618/machine-learning-when-to-use-each-method-and-technique.html> (дата обращения: 01.05.2021).