

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет
имени первого Президента России Б.Н. Ельцина»

Институт радиоэлектроники и информационных технологий - РТФ
Базовая кафедра «Аналитика больших данных и методы видеоанализа»

ДОПУСТИТЬ К ЗАЩИТЕ ПЕРЕД ГЭК
Зав. кафедрой АБДиМВ

«18» мая 2022 г.

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
(МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)**

«Разработка аналитического обеспечения технологии машинного обучения в
деятельности страховой компании»

Научный руководитель: Медведев М. А.

Научный руководитель: Коломыцева А.О.

Нормоконтролер: Медведева М.А.

Студент группы РИМ-201230 Денисенко Н. С.

Екатеринбург
2022

РЕФЕРАТ

Тема магистерской диссертации:

«Разработка аналитического обеспечения технологии машинного обучения в деятельности страховой компании»

Магистерская диссертация выполнена на 89 страницах, содержит 12 таблиц, 43 рисунков, 71 использованных источников и одно приложение.

Актуальность темы обусловлена значительным ростом использования информационных и цифровых решений, которые играют важную роль, поскольку являются важным ресурсом финансовой стабильности страховых компаний, которая может обеспечивать ее выигрышную позицию в конкурентной борьбе, учитывая усиление концентрации игроков страхового рынка.

Цель работы: повышение эффективности управления информационными ресурсами посредством разработки аналитического обеспечения технологии машинного обучения в деятельности страховой компании.

Для выполнения поставленной цели необходимо решение следующих задач:

- сформировать математическую постановку задачи кластеризации в алгоритмах машинного обучения;
- разработать алгоритм предикативной аналитики с использованием архитектурного подхода;
- осуществить анализ проблем функционирования и развития страхового рынка;
- определить задачу формирования базы данных о клиентах страховой компании;
- разработать модель полной архитектуры деятельности оператора страхового рынка;

- осуществить информационное обеспечение и поддержку технологии машинного обучения в деятельности оператора рынка страховых услуг;
- оценить влияние алгоритмов машинного обучения в страховании на эффективность работы оператора рынка;
- провести анализ экономического эффекта и рисков внедрения технологии машинного обучения в страховании.

Объектом исследования является управление информационными ресурсами страховой компании.

Предметом исследования являются методы разработки модели машинного обучения в сфере тарификации.

Научная новизна состоит в том, что построена полная модель архитектуры ПАО СК «Росгосстрах», разработана аналитическая модель машинного обучения, в основе которой заложен математический метод логистической регрессии для бинарной классификации, разработаны процессные модели реализации проекта по внедрению методов машинного обучения в деятельность страховой компании, разработана имитационная модель управления проектом по внедрению модели машинного в деятельность страховой компании.

Практическая значимость заключается в том, что разработанная модель машинного обучения повысит эффективность использования входных, а также уже имеющихся данных, ускорит процесс обработки информационных ресурсов и повысит эффективность деятельности оператора страховой компании в сфере тарификации.

Экономическая эффективность проекта по внедрения модели машинного обучения в деятельность страховой заключается в том, что применение данной модели увеличило среднюю страховую премию по ОСАГО, тем самым увеличив общую сумму страховых премий на 8,42 %. Использование данной модели способствует формированию индивидуальной страховой премии для каждого клиента и получению дополнительной прибыли для страховой компании.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	7
1 ТЕОРЕТИЧЕСКИЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ МАШИННОГО ОБУЧЕНИЯ В СТРАХОВАНИИ	13
1.1 Особенности и проблематика технологий машинного обучения в страховании.....	13
1.2 Математическая постановка задачи кластеризации в алгоритмах машинного обучения	20
1.3 Разработка алгоритма предикативной аналитики с использованием архитектурного подхода	26
1.4 Результаты и выводы	30
2 РАЗРАБОТКА АНАЛИТИЧЕСКИХ СИСТЕМ В СТРАХОВАНИИ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ.....	31
2.1 Анализ проблем функционирования и развития страхового рынка.....	31
2.2 Задача формирования базы данных о клиентах страховой компании	38
2.3 Разработка модели полной архитектуры деятельности оператора страхового рынка	49
2.4 Результаты и выводы	58
3 ОЦЕНКА РЕЗУЛЬТАТИВНОСТИ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ В СТРАХОВОЙ ДЕЯТЕЛЬНОСТИ	59
3.1 Информационное обеспечение и поддержка технологии машинного обучения в деятельности оператора рынка страховых услуг	59
3.2 Оценка влияния алгоритмов машинного обучения в страховании на эффективности работы операторов рынка	65
3.3 Оценка экономического эффекта и рисков внедрения технологий машинного обучения в страховании	71
3.4 Результаты и выводы	78
ЗАКЛЮЧЕНИЕ	79
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	81
ПРИЛОЖЕНИЕ А - Полная модель архитектуры ПАО СК «Росгосстрах»	89

ВВЕДЕНИЕ

Актуальность темы исследования. Тенденции развития постиндустриальной экономики ставят перед субъектами экономических отношений проблему сохранения накопленного потенциала, противодействия внешним угрозам и обеспечения глобальной конкурентоспособности. Находясь в мировом тренде неоиндустриального развития, Российская Федерация вынуждена решать две упомянутые проблемы развития синергетически: развиваться опережающими темпами и противостоять угрозам, сохраняя накопленный потенциал. Получение результатов в рамках такой комплексной и инновационно емкой стратегии должно сопровождаться «цифровым скачком» в парадигме управления, опережающими темпами цифровой трансформации на уровне взаимодействия агентов, сверхзадачами инновационно-технологического и кадрового обеспечения такой стратегии на уровне государственной политики.

Несмотря на все усилия в рамках стратегии цифровизации, Российская Федерация пока не входит в перечень стран-лидеров по уровню развития цифровой экономики, однако в данном направлении отмечается ряд положительных тенденций, которые подтверждают оценки Глобального института McKinsey [1]. Потенциальным эффектом от цифровизации экономики страны станет увеличение ВВП к 2025 году на 4,1–8,9 трлн. руб., что составит от 19–34% общего прогнозируемого роста показателя.

Основой таких оптимистичных прогнозов выступают ожидаемые эффекты от завершения автоматизации производственных процессов субъектов реального сектора, массовое внедрение современных информационных технологий в финансовом секторе и сфере оказания услуг, а также внедрение цифровых платформ, искусственного интеллекта, больших массивов данных (Big Data), роботизации, нейронных сетей и интернета вещей [2], как наиболее значимых цифровых проектов в сфере перехода к цифровому формату компаний.

Страхование является одним из важнейших элементов системы рыночных отношений и относится к финансовым отношениям, связанным с выполнением

специфических функций в экономике. Услуги страхования становятся все более популярными, ими пользуется значительная часть граждан РФ. Согласно данным Центробанка, по итогам 2021 г. в России работали 532 страховые компании. Услугами страхования пользуется определенная часть граждан Российской Федерации, формируя по различным оценкам емкость данного сегмента в пределах от 1,18 до 1,54 млрд. рублей по объемам начисленных премий [3]. Страхование является одним из важнейших элементов системы финансовых отношений, связанным с выполнением специфических финансовых функций. Объем произведенных выплат за 2020 г. составил 658,65 млрд. рублей. Самыми востребованными видами страхования оказались автострахование (354,39 млрд. рублей) и страхование жизни (333,70 млрд. рублей). Отталкиваясь от структуры страхового спроса, можно предположить, что сохранение внутренней эффективности и детальный анализ характеристик клиентов могут существенно повысить устойчивость работы его операторов.

Цифровая трансформация бизнеса подразумевает кардинальную смену экономического уклада за счёт интеграции современных цифровых технологий во все аспекты бизнес-деятельности. Переход к цифровой экономике меняет как общий уклад хозяйственной деятельности в обществе, так и модель деятельности отдельных субъектов экономического взаимодействия.

Современные информационные и цифровые решения играют важную роль, поскольку являются важным ресурсом финансовой стабильности страховых компаний, которая может обеспечивать ее выигрышную позицию в конкурентной борьбе, учитывая усиление концентрации игроков. Управление внутренними информационными ресурсами страховой компании базируется на использовании как традиционных методов и моделей, так и ряда перспективных инструментов, в основе которых лежат технологии управления данными. Для страховых компаний необходимо определять размер страховых платежей, опираясь на страховую историю прошлых клиентов. Реализовать данный подход к расчету стоимости страховых полисов можно с помощью использования инструментов машинного обучения.

Это отражает актуальность данной проблемы и тем самым обуславливает выбор темы исследования, его цель и задачи.

Степень разработанности проблемы. Проблемы управления информационными ресурсами нашли своё отражение в работах Година В. В. [4], Корнеева И. К. , Гринберга А. С. [5], Короля И. А., Кострова А. В. [6], Симонова Ю. Ф. [7], Бормотава В. В. Хорошилова А. В [8]., Авдонин Л. А. [9], Федулина А. А. [10]. Основной проблемой управления информационными ресурсами страховой компании является недостаточный уровень их использования. Решением данной проблемы является реализация проекта информатизации предприятия и переход к цифровому варианту.

Вопрос применения алгоритмов моделей машинного обучения рассматривался в работах зарубежных и отечественных исследователей, таких как П. Домингос [11], П. Флах [12] и Х. Бринк [13], Плоткина А.А. [14], Стефанова Н. А. [15], Массель Л.В. [16].

Цель и задачи исследования. Целью данной работы является повышение эффективности управления информационными ресурсами посредством разработки аналитического обеспечения технологии машинного обучения в деятельности страховой компании.

В соответствии с целью были поставлены следующие задачи:

изучить особенности и проблематику технологий машинного обучения в страховании;

сформировать математическую постановку задачи кластеризации в алгоритмах машинного обучения;

разработать алгоритм предикативной аналитики с использованием архитектурного подхода;

осуществить анализ проблем функционирования и развития страхового рынка;

определить задачу формирования базы данных о клиентах страховой компании;

разработать модель полной архитектуры деятельности оператора страхового рынка;

осуществить информационное обеспечение и поддержку технологии машинного обучения в деятельности оператора рынка страховых услуг;

оценить влияние алгоритмов машинного обучения в страховании на эффективность работы оператора рынка;

провести анализ экономического эффекта и рисков внедрения технологии машинного обучения в страховании.

Объект исследования. Объектом исследования является управление информационными ресурсами страховой компании.

Предмет исследования. Предметом исследования являются методы разработки модели машинного обучения в сфере тарификации.

Методы исследования. Методологическую основу магистерской диссертации составляют работы отечественных и зарубежных ученых в области: системного анализа (изучение теоретических и прикладных моделей управления информационными ресурсами страховой компании), архитектурного проектирования (разработка полной модели архитектуры деятельности ПАО СК «Росгосстрах»), процессного моделирования (разработка процессных моделей реализации этапов проекта по внедрению модели ML в деятельность страховой компании)), аналитическое моделирования (разработка алгоритма модели машинного обучения), системно-динамического моделирования (реализация имитационной модели управления проектом разработки и внедрения алгоритма модели ML), имитационного моделирования (проведение экспериментов в ходе реализации системно-динамической модели), проектного моделирования (экономическое обоснование проекта внедрения методов машинного обучения в деятельность страховой компании).

Научная новизна. В работе осуществлена постановка и решение новой актуальной задачи управление информационными ресурсами страховой компании. При этом были получены новые результаты:

получили дальнейшее развитие:

принципы построения полной модели архитектуры деятельности страховой компании, что обеспечивает целостный взгляд на деятельность предприятия, его ресурсы и возможности, а также повышение интегрированности предприятия и поддержку его анализа в самых различных разрезах (организационных, экономических, количественных, качественных и т.д.) для совершенствования деятельности по принятию решений, контролю, координации и мониторингу управленческих решений;

процедуры разработки аналитической модели машинного обучения, в основе которой заложен математический метод логистической регрессии для бинарной классификации. Данная модель позволяет проанализировать уже имеющиеся данные страховой компании и спрогнозировать вероятность наступления страхового случая нового клиента, тем самым повысит эффективность работы оператора страхового рынка.

Возможности использования имитационной модели управления этапами проекта (пространственно-временной аспект) по разработке и внедрению модели машинного в деятельность страховой компании, позволяющая определять длительность реализации проекта на основе различных сценариев.

Практическая значимость полученных результатов. Практическое значение полученных результатов состоит в том, что теоретические положения, изложенные в работе, доведены до уровня практических предложений и рекомендаций. Предложенная автором модель полной архитектуры деятельности предприятия, системно-динамическая модель управления проектом разработки и внедрения модели машинного обучения в деятельность страховой компании позволяют повысить эффективность использования информационных ресурсов. Разработанная модель машинного обучения повысит эффективность использования входных, а также уже имеющихся данных, ускорит процесс обработки информационных ресурсов и повысит эффективность деятельности оператора страховой компании в сфере тарификации.

Личный вклад автора. Магистерская диссертация является самостоятельно выполненным исследованием. Все результаты получены автором самостоятельно.

Апробация. Положения магистерской диссертации докладывались и обговаривались на VI Международной научно конференции «Донецкие чтения: образование, наука, инновации, культура и вызовы современности» (г. Донецк, 2021 г.), на II Международной научно-практической конференции студентов, аспирантов и молодых ученых «Инструменты проектного управления и анализ данных в системах поддержки принятия решений» (г. Донецк, 2021 г.), на I международной научно – практической конференции студентов, аспирантов и молодых ученых «Кибернетика, информатика аналитика: модели, инструменты, методы» (г. Донецк, 2022 г.), на III Международной научно-практической конференции студентов, аспирантов и молодых ученых «Инструменты проектного управления и анализ данных в системах поддержки принятия решений» (г. Донецк, 2022 г.).

Итоги работы представлены Международном конкурсе научно – исследовательский работ студентов, аспирантов и молодых ученых по экономическим наукам БелГУ им. Шухова (г. Губкин, 2021 г.) – диплом II степени.

Объем и структура работы. Магистерская диссертация состоит из введения, 3 разделов, заключения, выводов к разделам и списка литературы общим объемом 89 страниц, включая 43 рисунка, 12 таблиц, библиографического списка, состоящего из 71 источников и одного приложения.

1 ТЕОРЕТИЧЕСКИЕ АСПЕКТЫ ИСПОЛЬЗОВАНИЯ МАШИННОГО ОБУЧЕНИЯ В СТРАХОВАНИИ

1.1 Особенности и проблематика технологий машинного обучения в страховании

Область страхования РФ претерпевает стремительные изменения. Перспективы цифровизации страхового рынка были оценены страховщиками в ходе конъюнктурного исследования «Показатели цифровизации страхового рынка в Российской Федерации». Все страховые компании подтвердили намерение к внедрению новых цифровых технологий в ближайшие время (рис. 1). Наибольшей популярностью к внедрению у страховщиков пользуются следующие цифровые технологии:

- новые производственные технологии (телематика, телемедицина и прочие) (77 % страховщиков заявили о внедрении);
- технологии интернетизации бизнес-процессов (73 %);
- технологии больших баз данных (Big Data) (60 %);
- технологии искусственного интеллекта (47 %).

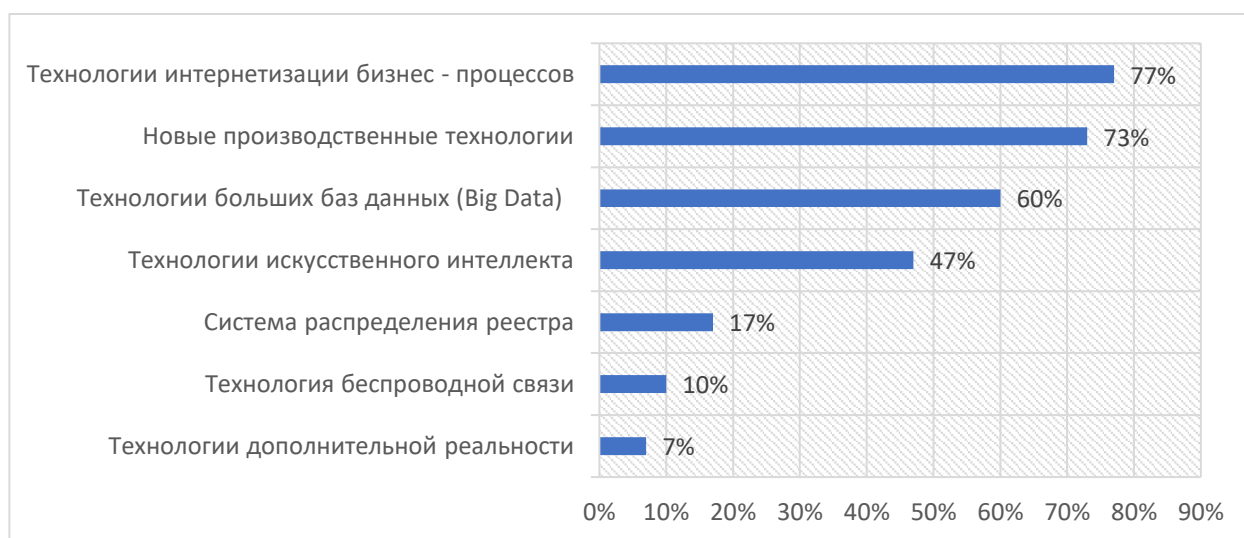


Рисунок 1 - Цифровые технологии, запланированные к внедрению страховыми компаниями в ближайшие время¹

¹ Составлено автором по: [17]

Использование технологии больших баз данных (Big Data) является основой использования моделей машинного обучения в деятельности страховой компании.

Процессы цифровизации должны учитываться органами регулирования страхового рынка, по-иному определяя доступность страховых услуг, конкурентную ситуацию на рынке, позволяя более точно определить страховой риск и рассчитать платежеспособность и финансовую устойчивость. Перспективные направления цифровизации на страховом рынке должны быть учтены ИТ-компаниями для разработки специализированных программ для страховых компаний.

Machine Learning (машинное обучение) — это множество математических, статистических и вычислительных методов для разработки алгоритмов, способных решить задачу не прямым способом, а на основе поиска закономерностей в разнообразных входных данных [18].

Сегодня машинное обучение эффективно используют для автоматизации задач, выполнение которых требует большого количества рутинного ручного труда и которые тяжело запрограммировать традиционным способом. Например, это задачи с большим количеством влияющих переменных: идентификация спам-писем, поиск информации в тексте и так далее. В таких ситуациях применение машинного обучения становится особенно востребованным.

Компании начинают активно инвестировать в машинное обучение, начинается полноценная гонка, в которой страховщики, адаптировавшиеся к условиям конкуренции, получают гораздо более качественные портфели договоров и довольных клиентов, уверенных в качестве сервиса. Сегодня машинное обучение — основной инструмент для реализации аналитических решений в страховании.

ML в страховании имеет свою практику применения. Новые методы анализа и появления дополнительных источников данных стали поводом переосмыслить роль математических методов в отрасли страхования [19]. Анализ опыта использования моделей машинного обучения в сфере страхования

позволяет установить будущие тенденции и проблемы внедренческого этапа. Оценка успешности внедрения и использования моделей ML в страховании в различных направлениях представлен на рисунке 2.

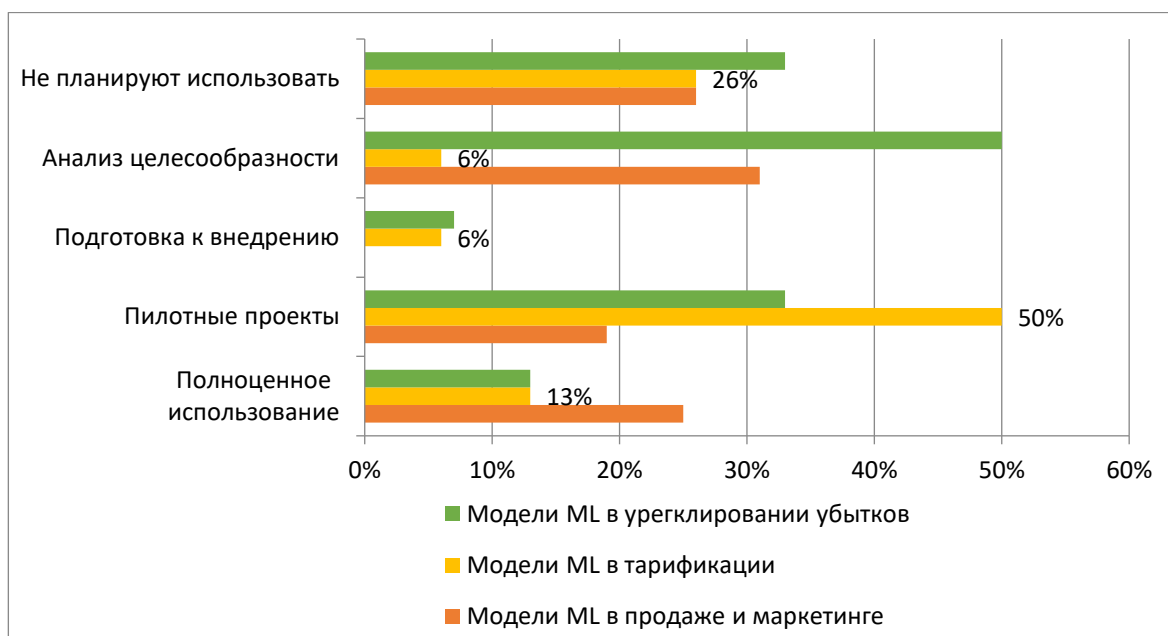


Рисунок 2 - Оценка статуса внедрения инструментов машинного обучения по итогам изучения существующих практик в 2020 году²

На рисунке 2 выделена сфера тарификации как наиболее значимая и учитывая высокий уровень пилотных проектов именно данная группа бизнес-процессов требует внимания аналитиков.

Цифровая трансформация бизнеса подразумевает кардинальную смену экономического уклада за счёт интеграции современных цифровых технологий во все аспекты бизнес-деятельности. Переход к цифровой экономике меняет как общий уклад хозяйственной деятельности в обществе, так и модель деятельности отдельных субъектов экономического взаимодействия. Важнейшей частью развития рынка страховых услуг и перехода к цифровому варианту являются информационные ресурсы. Они имеют огромное значение, поскольку являются важным ресурсом страховых компаний, который может обеспечивать ее выигрышную позицию в конкурентной борьбе [21].

² Составлено автором по: [20]

Применение новых информационных технологий становится принципиальным условием для достижения и удержания страховыми компаниями лидирующих позиций на рынке. Лидеры страхового рынка уже достигли высокого профессионализма непосредственно в области осуществления страховых операций, поэтому успешность страховщика во многом определяется технологической оснащенностью, так как от уровня технической подготовленности компании зависит скорость и качество обработки растущих потоков информации, что приводит к улучшению качества и повышению оперативности обслуживания клиентов [22].

Разработка проекта по цифровизации страховой компании является длительной и дорогостоящей задачей, поэтому предварительно должно быть проведено обоснование возможности и готовности компании работать в цифровом варианте (рис.3).

Этап	Сущность	Методы исследования
Предварительный	Постановка проблемы, анализ опыта успешных страховых компаний, описание недостатков в деятельности страховой компании, формулирование цели, задач исследования, формирование команды, определение структуры необходимых данных	SWOT – анализ, ПСА, методы бизнес-анализа, инструменты бенчмаркинга, организационное моделирование
Подготовительный	Диагностика деятельности страховой компании, анализ работы АИС, диагностика деятельности, сбор и анализ необходимых данных, построение и анализ процессной модели AS-IS, выявление узких мест и потерь.	Методы диагностики, бизнес – моделирование, логические и структурные методы
Аналитический	Поиск и обоснование направлений оптимизации бизнес-процессов страховой компании, построение процессной модели TO-BE, расчет стоимости бизнес- процесса, оценка затрат.	Бизнес-моделирование, расчетно-аналитический метод, реинжиниринг, методы и модели оптимизации бизнес-процессов.

Рисунок 3 - Схема процесса обоснования возможности перехода страховой компании к цифровому варианту ³

³ Составлено автором по: [23-24]

Информационные ресурсы представляют совокупность данных, организованных для получения достоверной информации в различных областях знаний и практической деятельности. Общим трендом последних лет в большинстве отраслей экономики, в том числе на рынке страхования является цифровизация, которая оказывает значительное влияние на деятельность страховщиков - хозяйствующих субъектов, осуществляющих страховую деятельность, являющихся основными участниками рынка страхования

На текущем этапе развития ML – это венчурная инвестиция. Как и любому венчурному предприятию, для успешного внедрения машинного обучения требуется специальная «среда» [25]. Это среда состоит из данных, специальных программных средств, культуры работы с данными, а главное, сотрудников с компетенциями в анализе данных и предметной области.

Выделяют следующие направления машинного обучения в страховании.

1. *Sales (продажи)*. Пролонгация и предсказание оттока клиентов.

Персонафицированное отношение к клиентам – направление, по которому идут многие компании, где-то, как в розничном банкинге, это уже является одним из важных конкурентных преимуществ. Для того чтобы решить успешно эту задачу, требуется понять поведение и потребности каждого конкретного клиента.

В этом нам помогают задачи сегментации клиентов по многим факторам: соцдем, информация о страховых продуктах, которыми клиент пользуется, история его обращений через службу поддержки и напрямую через филиалы компании и т. д. Это многофакторная модель с сотнями тысяч прецедентов и эффективно построить ее можно только силами высококвалифицированных аналитиков и экспертов с применением сложного математического аппарата, в том числе и с применением методов Машинного обучения.

После решения задач сегментации клиентов надо доработать полученную модель, находя группы лояльных и нелояльных клиентов (предсказание оттока), или находя группы людей, которые с большой долей вероятности готовы на покупку еще одного страхового продуктов (cross-sale или next-best-offer) [26].

2. *Underwriting* (андеррайтинг). Создание и оптимизация рисковых моделей, предсказание убыточности клиента.

Оценка рисков – крайне важный процесс для страхового бизнеса, требующий постоянного развития методов и инструментария.

Здесь различают два важных, дополняющих друг друга, направления.

Первое: обогащение имеющихся о клиенте данных (с его согласия), например, данными из БКИ.

Второе: применение интерпретируемых алгоритмов Машинного обучения, таких как линейные классификаторы, леса решений, метод главных компонент [27].

Сейчас применение дополнительных методов машинного обучения в андеррайтинге ограничено, причем не в силу технологической слабости, а наоборот. Классическая обобщенная линейная модель (GLM) при использовании мощных внешних баз дает очень неплохие результаты, которых другие решения пока не достигли.

Тем не менее, у новых ML методов есть своя ниша. Перерасчет ОЛМ и поиск новых факторов – это очень трудоемкая задача. Вместе с тем, другие интерпретируемые методы обрабатывают новые данные и обучаются быстрее, реагируя на изменения. Это помогает андеррайтерам как в поиске новых факторов, улучшающих рисковую модель, так и искать риски, недооцененные линейной моделью [28].

3. *Antifraud*. Борьба с мошенничеством.

Злонамеренное мошенничество, к сожалению, имеет место быть во многих областях. Борьба с ним часто неприятный, но необходимый процесс. Существует несколько способов машинного обучения. Такой способ машинного обучения, как обучение с частичным привлечением учителя, можно применять для распознавания мошенничеств с попытками выдать себя за иного клиента. Мошенничества можно классифицировать как аномалию на фоне обычной активности. Методы машинного обучения с частичным привлечением учителя позволяют создавать модели, которые распознают такие аномалии.

Соответствующие системы нередко применяются для выявления попыток мошенничеств при онлайн - сделках.

Для эффективного мониторинга все новых и новых мошеннических схем требуется отслеживать и обрабатывать все большее количество данных. Как и в случае со скоринговыми моделями, автоматизация и ML помогает поддерживать нужную скорость процесса обнаружения мошенничества и в автоматическом режиме, за сотые секунды оценивать вероятность мошенничества по каждому конкретному страховому договору. Такая система применяется в Турции, где работа с данными по мошенникам централизована на базе единого бюро страховых историй.

Кроме того, есть и кросс-секторальные ML - кейсы, которые могут быть использованы в страховании. Например, семантический анализ информационного поля (СМИ, публичной части социальных сетей) с целью превентивного обнаружения информационных атак, анализа отношения к бренду или оптимизации маркетинговых кампаний. Эти задачи могут быть решены с помощью Natural Language Processing – подраздел машинного обучения, занимающийся лингвистическими задачами [29].

Тем не менее, на российском рынке убытки по мошенничеству в каждой отдельно взятой компании еще не сформировали достаточного массива данных для построения модели способной самостоятельно предсказать мошенничество с приемлемой точностью. До момента запуска закона о бюро страховых историй и технического перевооружения ВСС в этой области оперативное чутье скорее всего будет существенно превосходить искусственный интеллект. Текущие технические решения в основном помогают фильтровать небольшие убытки перед ручной проверкой по «красным флагам» [30].

Страховые компании внедряют системы, которые используют облачные хранилища, что позволяет уменьшить физический контакт между клиентами и страховщиками, а также ускорить процесс передачи и обработки данных между отделами компании. Внедрение когнитивных технологий и технологий искусственного интеллекта обеспечит сбор больших объемов данных с помощью

устройств Интернета вещей, последующий анализ которых позволит улучшить скоринговые модели, с помощью которых производится оценка возможности будущей убыточности владельца страхового полиса на основании данных из его кредитной истории.

1.2 Математическая постановка задачи кластеризации в алгоритмах машинного обучения

Общий термин «Machine Learning» или «машинное обучение» обозначает множество математических, статистических и вычислительных методов для разработки алгоритмов, способных решить задачу не прямым способом, а на основе поиска закономерностей в разнообразных входных данных. Решение вычисляется не по четкой формуле, а по установленной зависимости результатов от конкретного набора признаков и их значений [31].

Поэтому машинное обучение применяется для диагностики, прогнозирования, распознавания и принятия решений в различных прикладных сферах: от медицины до банковской деятельности.

Выделяют 2 типа машинного обучения:

1. Индуктивное или по прецедентам, которое основано на выявлении эмпирических закономерностей во входных данных;
2. Дедуктивное, которое предполагает формализацию знаний экспертов и их перенос в цифровую форму в виде базы знаний [27].

Дедуктивный тип принято относить к области экспертных систем, поэтому общий термин «машинное обучение» означает «обучение по прецедентам».

Прецеденты или обучающая выборка – это наборы входных объектов и соответствующих им результатов. При этом не существует четкой формулы, которая аналитически описывает зависимость между результатами и входами.

Можно сказать, что машинное обучение реализует подход Case Based Reasoning — метод решения проблем рассуждением по аналогии, путем предположения на основе подобных случаев (прецедентов).

Существует 5 способов машинного обучения [32].

1) Обучение с учителем (Supervised Learning), когда необходимо найти функциональную зависимость результатов от входов и построить алгоритм, на входе принимающий описание объекта и на выходе выдающий ответ. Функционал качества, как правило, определяется через среднюю ошибку ответов алгоритма по всем объектам выборки. К обучению с учителем относятся задачи классификации, регрессии, ранжирования и прогнозирования.

2) Обучение без учителя (Unsupervised Learning), когда ответы не задаются, и нужно искать зависимости между объектами. Сюда входят задачи кластеризации, поиска ассоциативных правил, фильтрации выбросов, построения доверительной области, сокращения размерности и заполнения пропущенных значений.

3) 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 остальное.

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 [33].

4) 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.

Роботы, используя обучение с подкреплением и глубинное обучение, определяют, насколько сильно нужно сдавить предмет при хватании, и какой хват будет наилучшим. Разновидность этого способа, глубинное обучение с подкреплением, хорошо подходит для автономного принятия решений в случаях, когда возможностей обучения с учителем и без недостаточно.

5) Глубинное обучение (Deep learning). Глубинное обучение может проходить как без учителя, так и с подкреплением. При глубинном обучении частично имитируются принципы обучения людей — 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 тоже широко применяется для борьбы с мошенничествами, так как улучшает точность распознавания благодаря автоматизации. Сегодня страховщики осознают возможности современных информационных технологий в раскрытии и предупреждении подобных преступлений, увеличивая бюджеты на внедрение Big Data систем предиктивной и графовой аналитики с помощью алгоритмов Machine Learning [34].

Наиболее востребованными в сфере страхования являются следующие технологии:

- предикативное моделирование;
- анализ взаимодействия юридических и физических лиц, включая анализ социальных сетей;
- глубинный анализ текста;
- обнаружение аномалий.

В основе машинного обучения лежат математические и аналитические методы, которые способны решить задачу путем поиска закономерностей в разнообразных входных данных [35].

В таблице 1 представлены методы машинного обучения в сферах страхования: управление продажами и продвижение страховых продуктов, тарификация страховых премий и выплат (скоринговые модели), урегулирование рисков, убытков и предсказание потерь.

Таблица 1 - Систематизация базовых алгоритмов обучения модели управления данными с учетом специфики бизнес-процессов страховой компании⁴

Бизнес-процессы как условия формирования данных	Алгоритмы машинного обучения с учетом системы признаков	Инструменты машинного обучения
Управление продажами и продвижение страховых продуктов	линейная регрессия*, (для числовых данных), линейный дискриминантный анализ (для мультикатегорийной классификации);	обучение целевой функции f , которая самым лучшим образом соотносит входные переменные x и выходную переменную y : $y = f(x)$. Отличия в характеристиках данных, их объемах и требованиях, числе классов.
	кластеризация методом k -ближайших соседей	
Тарификация страховых премий и выплат (скоринговые модели)	логистическая регрессия* (для бинарной классификации);	обучение в рамках алгоритма различными требованиями к объему хранения данных; векторы выбираются случайным образом (Azure ML, система RapidMiner, Knime, IBM SPSS Modeler)
	обучение нейронной сети Кохонена	
	LVQ — сети векторного квантования	
Урегулирование рисков, убытков и предсказание потерь	бэггинг, бутстрэп, деревья решений, ассоциативные правила и генетические алгоритмы**;	для всех выборок из тренировочных данных строят деревья решений в соответствии с ассоциативными правилами (IBM Watson Analytics, SAS Enterprise Miner)
	методы бустинга, включая AdaBoost и XGBoost, являются ансамблями алгоритмов	создают серию моделей, где каждая новая модель пытается исправить ошибки предыдущей (SAP BusinessObjects Predictive Analytics, Oracle Big Data Preparation)

Логистическая регрессия — это разновидность множественной регрессии, общее назначение которой состоит в анализе связи между несколькими независимыми переменными (называемыми также регрессорами или предикторами) и зависимой переменной. Бинарная логистическая регрессия применяется в случае, когда зависимая переменная является бинарной (т.е. может принимать только два значения) [37]. С помощью логистической

⁴ Составлено автором по: [36-37]

регрессии можно оценивать вероятность того, что событие наступит для конкретного испытуемого. Все регрессионные модели могут быть записаны в виде формулы:

$$y = F(x_1, x_2 \dots x_n)$$

В множественной линейной регрессии предполагается, что зависимая переменная является линейной функцией независимых переменных, т. е.:

$$y = a + b_1x_1 + b_2x_2 + \dots b_nx_n$$

Данная модель используется для задачи оценки вероятности исхода события путём вычисления стандартных коэффициентов регрессии. Например, если рассматривается исход по наступления страхового случая, задается переменная y со значениями 1 и 0, где 1 означает, что с соответствующим клиентом произойдет страховой случай, а 0, что страхового случая не произойдет [38].

Однако здесь возникает проблема: множественная регрессия не «знает», что переменная отклика бинарная по своей природе. Это неизбежно приведет к модели с предсказываемыми значениями большими 1 и меньшими 0. Но такие значения вообще не допустимы для первоначальной задачи. Таким образом, логистическая регрессия просто игнорирует ограничения на диапазон значений для y .

Для решения проблемы задача регрессии может быть сформулирована иначе: вместо предсказания бинарной переменной, мы предсказываем непрерывную переменную со значениями на отрезке $[0,1]$ при любых значениях независимых переменных. Это достигается применением следующего регрессионного уравнения (логит-преобразование):

$$P = \frac{1}{1 + e^{-y}}$$

где P — вероятность того, что произойдет интересное событие; e — основание натуральных логарифмов 2,71...; y — стандартное уравнение регрессии [39].

ROC-кривая (Receiver Operator Characteristic) — кривая, которая наиболее часто используется для представления результатов бинарной классификации в машинном обучении. Название пришло из систем обработки сигналов. Поскольку классов два, один из них называется классом с положительными исходами, второй — с отрицательными исходами. ROC-кривая показывает зависимость количества верно классифицированных положительных примеров от количества неверно классифицированных отрицательных примеров.

В терминологии ROC - анализа первые называются истинно положительным, вторые — ложно отрицательным множеством. При этом предполагается, что у классификатора имеется некоторый параметр, варьируя который, мы будем получать то или иное разбиение на два класса. Этот параметр часто называют порогом, или точкой отсечения. В зависимости от него будут получаться различные величины ошибок I и II рода. В логистической регрессии порог отсечения изменяется от 0 до 1 — это и есть расчетное значение уравнения регрессии. Будем называть его рейтингом [40].

Для понимания сути ошибок I и II рода формируется четырехпольная таблица сопряженности (табл. 2), которая строится на основе результатов классификации моделью и фактической (объективной) принадлежностью примеров к классам.

Таблица 2 -Четырехпольная таблица сопряженности логистической регрессии [41]

	Фактически положительно	Фактически отрицательно
Положительно	TP	FP
Отрицательно	FN	TN

- TP (*True Positives*) — верно классифицированные положительные примеры (так называемые истинно положительные случаи).

- TN (*True Negatives*) — верно классифицированные отрицательные примеры (истинно отрицательные случаи).

- FN (*False Negatives*) — положительные примеры, классифицированные как отрицательные (ошибка I рода). Это так называемый «ложный пропуск» — когда интересующее нас событие ошибочно не обнаруживается (ложно отрицательные примеры).

- FP (*False Positives*) — отрицательные примеры, классифицированные как положительные (ошибка II рода). Это ложное обнаружение, т. к. при отсутствии события ошибочно выносится решение о его присутствии (ложно положительные случаи).

Машинное обучение имеет свои сферы применения, учитывая особенности алгоритмов, которые используются в отдельных сегментах оказания страховых услуг [42].

1.3 Разработка алгоритма предикативной аналитики с использованием архитектурного подхода

Машинное обучение в последнее время стало популярным трендом, но в целом его эффективность несколько переоценена. ML сейчас находится на этапе переосмысления в страховом секторе. Незрелость ML в отрасли часто приводит к завышенным ожиданиям, которые часто сменяются разочарованием руководства и специалистов страховщика в потенциале машинного обучения для бизнеса.

Машинное обучение включает в себя множество подходов и методов, в том числе нейросети, линейную регрессию, бустинги, решающие деревья, случайные леса и кластеризация. Наилучшим методом машинного обучения для страховой компании в сфере тарификации является бинарная логистическая регрессия, которая входит в предикативную аналитику.

Такой метод машинного обучения, как предикативное моделирование и аналитика используется в страховании для определения вероятности

наступления страхового случая с тем или иным клиентом на основе страховых историй прошлых клиентов [43].

Предикативная аналитика (Predictive analytics) — это множество методов статистики и анализа данных, которые используются для анализа текущих и исторических данных/событий для прогноза данных/событий в будущем. Наиболее известный способ использования прогностической аналитики — это применение скоринговые моделей для оценки платежеспособности клиента при выдаче кредитов в банке. Однако это не все области, где применяется предикативная аналитика, ее можно применять для разработки продуктов, для выбора потенциальной аудитории, для выбора следующего продукта, который можно предложить клиенту и множестве других.

Предиктивная аналитика использует разнообразный набор расчетных инструментов: статистические методы; теорию игр; интеллектуальный анализ данных, анализ текущих и исторических событий для составления прогноза. Посредством предиктивной аналитики можно спрогнозировать множество вариантов исхода события или поведения клиента. В результате, в наличии уже будет несколько вариантов реагирования, что в меняющихся условиях позволит быстро принять оптимальное решение. Этот метод аналитики основан на обработке больших массивов исторических данных [44].

Также, вместе с бинарной логистической регрессией используется метод «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, и так до тех пор, пока не будет достигнут «лист» — окончательный ответ. Среди применений дерева решений — платформы управления знаниями для клиентского обслуживания, прогнозного назначения цен 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

поданных требований и затраченных сумм. Затем, используя модели, система будет оценивать новые заявления о страховой защите, классифицируя их по категории риска и возможному финансовому ущербу [45].

Большинство актуариев используют обобщенную линейную модель для анализа тарифных факторов, а она построена на методе линейной регрессии, который входит в область машинного обучения. У методов машинного обучения есть много признаков группировки, но для практических целей ML-методы обычно группируются следующим образом:

- во-первых, по возможностям интерпретации решения. Какие-то методы четко оценивают влияние каждого фактора на итоговое решение, а другие выдают результат без оценки факторов влияния. Как правило, последние дают более точное решение. Для методов распознавания образов критерии принятия решения не так важны, а для тарифа наоборот;

- во-вторых, методы отличаются по пригодности для решения различных типов задач с данными различной структуры. Задачи по распознаванию изображения решаются одними алгоритмами, а другие помогают найти правильный тариф. Иногда методы комбинируются. Страховщики используют, как правило, случайный лес и прочие деревья.

Основу прогнозной аналитики составляет автоматический поиск закономерностей, взаимосвязей, аномалий между различными факторами. Как работает прогнозное моделирование: осуществляется построение модели на основе выявленных закономерностей при анализе исторических событий; основываясь на выявленных закономерностях осуществляется прогноз будущих процессов; на основании сделанного прогноза проводятся мероприятия, направленные на оптимизацию бизнес-процессов.

Центральной же сущностью предиктивной аналитики является задача определение предиктора или нескольких предикторов (параметров или сущностей, которые влияют на прогнозируемое событие) [46].

Страховые компании в сфере тарификации выделяют такие предикторы при определении суммы страховой премии: *возраст, пол, наличие детей, стаж*

возждения, тип автомобиля [47]. Множество этих предикторов образует модель предиктивной аналитики на основе прошлых данных о клиентах, которая предсказывает определенное событие в будущем с какой-то степенью вероятности. На основе такого прогнозирования принимаются решения. Модель машинного обучения на основе метода логистической регрессии определяет вероятность наступления страхового случая у того или иного клиента, что поспособствует предоставлению компанией персонального пакета услуг.

В процессе разработки моделей машинного обучения часто актуальна реализация программного комплекса в виде веб-сервиса. Для реализации машинного обучения рассматриваются 2 наиболее известных решения в этой области: MS Azure ML и Amazon ML [48].

Так как Amazon ML поддерживает работу только 1 алгоритма — линейную регрессию (и его адаптацию для задач классификации — логистическую), то это ограничивает возможности реализуемых решений. Microsoft Azure ML представляет собой более гибкий сервис:

- большое количество встроенных алгоритмов и поддержка встраивания своего кода на R и Python;
- во встроенных инструментах есть все необходимое для базовой работы в областях классификации, кластеризации, регрессии, компьютерного зрения, работы с текстом и др.;
- есть функционал для предобработки и манипуляции данными;
- модули в Azure ML организованы как блок-схемы, что делает порог вхождения низким, а работу интуитивной — поэтому Azure ML стал удобным средством прототипирования, позволяющим быстро реализовать базовые решения для проверки жизнеспособности гипотез, идей и проектов [49].

Деятельность достаточно большого количества страховых компаний связана с рисками. Предиктивный анализ позволяет обработать прошлые негативные события, чтобы предотвратить риск их возникновения в будущем или составить эффективный план по выходу из ситуации.

1. 4 Результаты и выводы

На основании анализа научных работ изучена технология использования машинного обучения в деятельности страховой компании. Рассмотрены основные направления и методы машинного обучения, которые могут быть использованы в различных сферах страховых услуг. Оценив статус внедрения инструментов машинного обучения по итогам изучения существующих практик в 2020 году, сфера тарификации выделена как наиболее значимая и учитывая высокий уровень пилотных проектов именно данная группа бизнес-процессов требует внимания аналитиков.

Проведена систематизация базовых алгоритмов обучения модели управления данными с учетом специфики бизнес-процессов страховой компании. Сформирована математическая постановка задачи кластеризации в алгоритмах машинного обучения. Предоставлено математическое описание логистической регрессии для бинарной классификации, как метода машинного обучения в сфере тарификации страховой компании.

Детально рассмотрен такой метод машинного обучения, как предикативное моделирование. Данный метод используется в страховании для определения вероятности наступления страхового случая с тем или иным клиентом на основе страховых историй прошлых клиентов. Рассмотрены два наиболее известных веб-сервиса для реализации проекта машинного обучения. Определены преимущества MS Azure ML в данной сфере. Данная платформа будет использована в дальнейшем для разработки алгоритма модели машинного обучения в сфере тарификации.

Внедрение алгоритма машинного обучения в деятельность страховой компании повысит эффективность использования входных, а также уже имеющих данных, ускорит процесс обработки информационных ресурсов и повысит эффективность деятельности оператора страховой компании в сфере тарификации.

2 РАЗРАБОТКА АНАЛИТИЧЕСКИХ СИСТЕМ В СТРАХОВАНИИ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

2.1 Анализ проблем функционирования и развития страхового рынка

Услугами страхования пользуется определенная часть граждан Российской Федерации, формируя по различным оценкам емкость данного сегмента в пределах от 1,18 (2018 г.) до 1,54 (2020 г.) млрд. рублей по объемам начисленных премий. Страхование является одним из важнейших элементов системы финансовых отношений, связанным с выполнением специфических финансовых функций. Объем произведенных выплат за 2020 г. составил 658,65 млрд. рублей (против 610,74 млрд. рублей в 2019 г.). Самыми востребованными видами страхования оказались автострахование (354,39 млрд. рублей) и страхование жизни (333,70 млрд. рублей). Динамика рынка страховых услуг представлена на рис. 4.

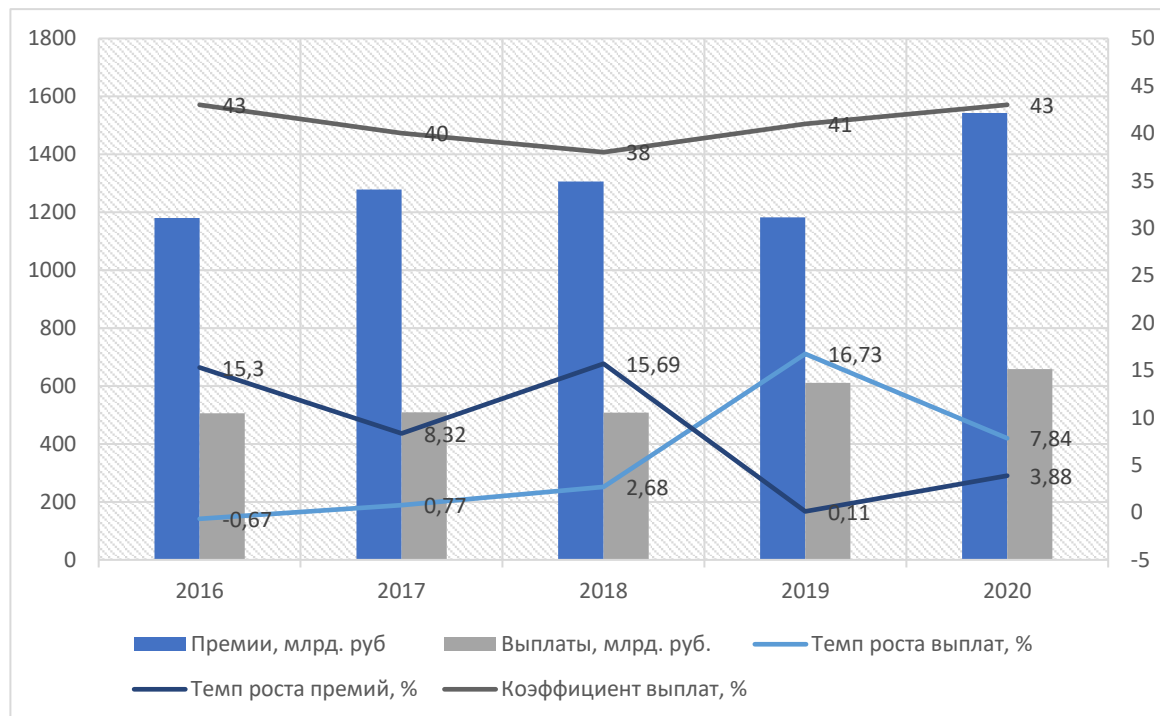


Рисунок 4 - Динамика показателей функционирования страхового рынка [50]

Отталкиваясь от структуры страхового спроса, можно предположить, что сохранение внутренней эффективности и детальный анализ характеристик клиентов могут существенно повысить устойчивость работы его операторов.

Наибольшую востребованность у страховщиков показывают технологии «интернетизации» бизнес-процессов» (93 % респондентов), технологии больших баз данных (57 %) и новые производственные технологии (47 %). В меньшей степени применяются технологии беспроводной связи, которые используют 40 % страховщиков, и технологии искусственного интеллекта (30 %). Страховщики практически не используют технологии виртуальной и дополненной реальности, блокчейн-технологии (рис. 5).

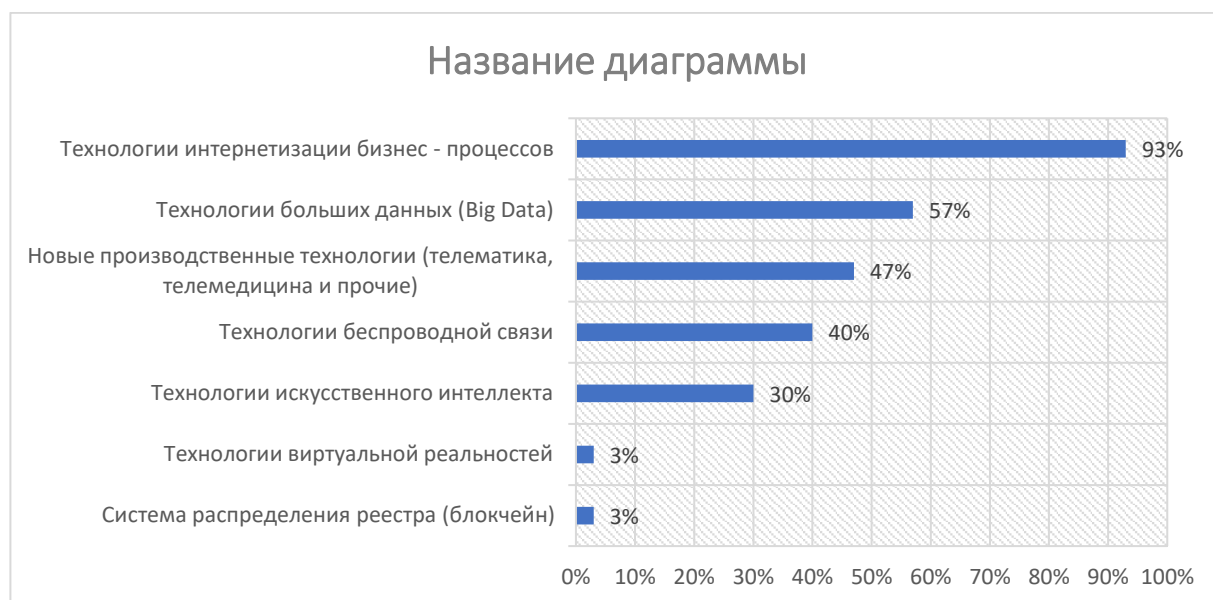


Рисунок 5 – Использование страховщиками новые цифровые технологии [51]

Во многом уровень цифровизации и использование отдельных цифровых технологий определяется бизнес-процессами в страховых компаниях. Наибольший показатель использования цифровых технологий (97 %) отмечается в бизнес-процессах продаж программ страхования, где преимущественно используются технологии производственного интернета (организация интернет-продаж полного цикла).

Технологии мобильных приложений в продажах используют 11 % страховщиков, но практика организации венчурных студенческих проектов и хакатонов в 2019–2020 гг., в которых принимали участие студенты Департамента страхования и экономики социальной сферы Финансового университета при Правительстве России, показывает, что интерес у крупных страховых компаний к данной технологии растет.

Цифровизация коммуникаций со страхователями / застрахованными лицами реализуется через технологии интернет-систем — организация личных кабинетов, чатов на сайтах и в мобильных приложениях (90 %), а также через электронный документооборот (57 %). В изучаемом периоде российские страховщики проявляли повышенную активность по внедрению новых цифровых технологий в бизнес-процессы урегулирования страховых случаев и выплат по договорам страхования, о чем заявили 80 % страховщиков. Наиболее часто используемой технологией выступает электронный документооборот (63 %), заключающийся в возможности предоставления заявления и документов на выплаты в электронном виде. Новая производственная технология в виде получения информации о страховом случае, принятие решения о выплате и осуществление выплаты в автоматическом режиме свойственна 37 % страховщиков [52].

Использование 30 % страховщиков технологий больших баз данных и 20 % страховщиков — технологий искусственного интеллекта (в основном обучаемых моделей) обусловлена необходимостью выявления и отсекания случаев с признаками страхового мошенничества. По личным видам страхования, особенно ДМС, 27 % страховщиков применяют новые производственные технологии в виде телемедицины (табл. 3).

Оценка страхового риска на этапе заключения договора осуществляется страховыми компаниями с использованием технологий больших баз данных (50 %) и искусственного интеллекта (20 %) для выявления и анализа математических закономерностей; применяются также новые производственные технологии (телематика, телездоровье и прочие) для оценки индивидуальных показателей по

объектам страхования (30 %). В бизнес-процессах управления рисками страховых компаний наибольшее распространение получили технологии системы больших баз данных (57 %) и искусственный интеллект (23 %).

Таблица 3 - Приоритетные цифровые технологии по бизнес-процессам в страховых компаниях⁵

Бизнес-процесс	Приоритетные цифровые технологии, % страховщиков, их применяющих
Продажи страховых программ	Технологии производственного интернета (97 %) Мобильные приложения (11 %)
Разработка программ страхования	Технологии производственного интернета (63 %) Новые производственные технологии, такие как телематика или телемедицина (43 %)
Коммуникации со страхователями	Технологии производственного интернета (организация личных кабинетов, чатов на сайтах и в мобильных приложениях) (90 %) Новые производственные технологии (электронный документооборот) (57 %)
Урегулирование страховых случаев и выплаты	Новые производственные технологии (электронный документооборот) (63 %) Новые производственные технологии (автоматизация выплат) (37 %) Технологии больших баз данных (30 %) Новые производственные технологии в виде телемедицины (27 %)
Оценка страхового риска	Технологии больших баз данных (50 %) Новые производственные технологии (телематика, телездоровье и прочие) (30 %) Искусственный интеллект (20 %)
Управление рисками	Технологии системы больших баз данных (57 %) Искусственный интеллект (23 %)

Несмотря на достаточно высокие показатели использования на страховом рынке новых цифровых технологий, страховые компании отмечают наличие

⁵ Составлено автором по: [53]

факторов, сдерживающих проникновение цифровизации в бизнес-процессы (рис. 6).



Рисунок 6 - Ключевые причины недостаточного проникновения цифровых технологий на рынок страхования [53]

В качестве ключевых причин недостаточного проникновения цифровых технологий в деятельность страховых компаний выделяются:

- высокая стоимость новых цифровых технологий (50 % страховщиков);
- отсутствие спроса на страховые программы с применением цифровых технологий (47 %);
- недостаточная адаптация цифровых технологий под страховые программы (47 %);
- недостаток квалифицированных кадров в области новых цифровых технологий (43 %).

Во многом источник причин недостаточного проникновения новых цифровых технологий следует искать вне страхового рынка. Отсутствие спроса на оцифрованные страховые программы может быть связано не только с низким уровнем проникновения страхования в экономику нашей страны, но и с низким уровнем цифровизации отраслей российской экономики, по мере повышения которого можно ожидать увеличение спроса на страховые программы,

предлагаемые в интернете [54]. Недостаточная адаптация цифровых технологий для страховых услуг в определенной степени обусловлена недостатком квалифицированных кадров в области новых цифровых технологий и страхового бизнеса. Важнейшим направлением снижения стоимости новых цифровых технологий является разработка универсальных (коробочных) решений цифровых технологий, что вызовет спрос существенного числа страховых компаний.

Цифровая экономика не изменяет экономической сути страхования как экономических отношений, связанных с наличием страхового риска и с формированием страхового фонда. Особенность страхования в цифровой экономике определяется: — специфическими рисками, присущими цифровой экономике и формирующими потребность в страховой защите; — специфическими цифровыми технологиями, составляющими основу цифровой экономики, и их применением на страховом рынке [55]. На рисунке 7 приводится схематичное определение направлений цифровизации страхования и их взаимосвязей в условиях цифровой экономики.



Рисунок 7– Направления цифровизации страхования [55]

Цифровое страхование выступает механизмом удовлетворения традиционной или специфической, возникшей в условиях цифровой экономики потребности в страховой защите посредством цифровых технологий. Деятельность страховых компаний с применением новых цифровых технологий трактуется как цифровизация страхового рынка.

Развитие цифрового страхования в части страхования специфических рисков цифровой экономики будет определяться:

- уровнем внедрения цифровых технологий в российскую экономику;
- развитием и сложностью цифровых технологий, прежде всего искусственного интеллекта и интернета вещей, в том числе интернета вещей в промышленности;
- готовностью российского законодательства по вопросам страхования киберрисков и других рисков цифровой экономики;
- развитием инфраструктуры страхования цифровых рисков, направленной прежде всего на предстраховой аудит и оценку причин наступления страхового случая;
- уровнем систем защиты от реализации цифровых рисков [56].

Однако при реализации проекта информатизации компании и использовании новых технологий, компании могут столкнуться со следующими препятствиями:

- 1) Ограничивающая регуляторная среда;
- 2) Неготовность потребителя страховых услуг;
- 3) Недостаток готовых ИТ-решений;
- 4) Отсутствие четкого понимания выгоды от внедрения новых технологий;
- 5) Недостаток необходимых данных [57].

На текущий момент основным барьером на пути развития новых технологий страховщики считают ограничивающую регуляторную среду. И, напротив, недостаток квалифицированных кадров перестал быть серьезным препятствием для менеджмента.

2.2 Задача формирования базы данных о клиентах страховой компании

Для оценивания вероятности наступления страхового случая у того или иного клиента будет использована онлайн-платформа Microsoft Azure ML [58]. Используемые данные — это данные страховой компании, которые включают информацию о прошлых клиентах. Компания предлагает широкий спектр продуктов, большая часть которых ориентирована на страхование автомобильного транспорта.

Полученные данные от страховой компании конвертируются в формат «.csv» и импортируются в Microsoft Azure ML, тем самым образуя «dataset» для дальнейшего анализа. Часть загруженной базы данных представлена на рисунке 8.

id	Name	№ Treaty	Sex	Age	Children	Class auto	Insurance case
1	Braund	4792	Женский	45	0	C,D	0
2	Cumings	4908	Мужской	57	0	A,B	0
3	Heikkinen	3635	Мужской	37	1	A,B	0
4	Futrelle	1178	Мужской	31	1	C,D	0
5	Allen	4989	Женский	53	0	A,B	0
6	Moran	2806	Женский	37	1	E,S	0
7	McCarthy	4112	Мужской	27	1	A,B	0
8	Palsson	3855	Мужской	64	0	C,D	0
9	Johnson	1880	Мужской	18	0	E,S	1
10	Nasser	3161	Мужской	20	1	C,D	0

Рисунок 8 - База данных клиентов страховой компании⁶

База данных, используемая в разработке алгоритма и модели, содержит 880 записей (клиентов) и 7 основных признаков:

- *Name* – Имя клиента.
- *№ Treaty* – Номер договора страхования по ОСАГО.

⁶ Составлено автором на основе полученных данных

- *Sex* – пол клиента.
- *Age* – возраст клиента.
- *Children* – наличие детей у клиента (1 – есть дети, 0 – нет детей).
- *Class auto* – класс автомобиля клиента.
- *Insurance case* – страховой случай с данным клиентом (1 – страховой случай зафиксирован, 0 – страхового случая не было).

Для изучения и анализа данных в будущем, необходимо исключить данные, не влияющие на наступление страхового случая (*id*, имя клиента, номер договора страхования). После выбора необходимых признаков база данных имеет следующий вид (рис. 9).

view as  

	Sex	Age	Children	Class auto	Insurance case
Женский	45	0	C,D	0	
Мужской	57	0	A,B	0	
Мужской	37	1	A,B	0	
Мужской	31	1	C,D	0	
Женский	53	0	A,B	0	
Женский	37	1	E,S	0	
Мужской	27	1	A,B	0	
Мужской	64	0	C,D	0	
Мужской	18	0	E,S	1	

Рисунок 9 – База данных после исключения не влияющих признаков⁷

База данных имеет не только числовые признаки, но и текстовые. Для того, чтобы текстовые признаки (*Sex*, *Class auto*) использовались при анализе данных, необходимо их преобразовать в категориальные переменные. База данных с преобразованными признаками представлена на рисунке 10. Полученные данные для дальнейшей разработки модели будут разделены на обучающие данные и тестируемые данные – 660 и 220 записей соответственно.

⁷ Составлено автором в Microsoft Azure ML

Sex-Женский	Sex-Мужской	Age	Children	Class auto-A,B	Class auto-C,D	Class auto-E,S	Insurance case
1	0	45	0	0	1	0	0
0	1	57	0	1	0	0	0
0	1	37	1	1	0	0	0
0	1	31	1	0	1	0	0
1	0	53	0	1	0	0	0
1	0	37	1	0	0	1	0
0	1	27	1	1	0	0	0
0	1	64	0	0	1	0	0
0	1	18	0	0	0	1	1

Рисунок 10 – Преобразованные данные⁸

В процессе обучения преобразованных данных будет использоваться метод «*Two – class logistic regression*». Логистическая регрессия - известный в статистике метод, который используется для прогнозирования вероятности исхода и особенно популярен для задач классификации. Алгоритм предсказывает вероятность возникновения события, подгоняя данные к логистической функции. В этом модуле алгоритм классификации оптимизирован для бинарных переменных. Параметры логистической регрессии представлены на рисунке 11.

Two-Class Logistic Regression

Create trainer mode
Single Parameter

Optimization tolerance
1E-07

L1 regularization weight
1

L2 regularization weight
1

Memory size for L-BFGS
20

Random number seed

Рисунок 11 – Параметры «Two – class logistic regression»⁹

⁸ Составлено автором в Microsoft Azure ML

⁹ Составлено автором в Microsoft Azure ML

Optimization tolerance - граница для минимального изменения точности. Если разница между итерациями становится меньше указанного значения, обучение останавливается. В данном случае значение по умолчанию равно $1E-7$.

L1 & L2 regularization — регуляризация по нормам 1 и 2 порядка. L1 используется для уменьшения размерности входных данных через удаление признаков, наименее влияющих на результат. L2 «штрафует» признаки с большими коэффициентами, сводя эти коэффициенты к нулю на бесконечности. Указанные значения являются коэффициентами для норм и для начала могут по умолчанию оставаться равными 1.

L-BFGS — метод оптимизации для нахождения локального максимума (минимума) нелинейного функционала без ограничений с ограниченной памятью. Это квазиньютоновский метод, используемый для взвешивания параметров модели. Увеличение этого значения улучшает точность, при этом может сильно увеличить время выполнения, поэтому часто выставляется в окрестности 10 [59].

Применив данный метод к обучаемым данным, определились наиболее значимые признаки в наступление страхового случая прошлых клиентов компании (рис. 12)

Feature Weights

Feature	Weight
Age	-4.08594
Class auto-E,S	2.66314
Class auto-A,B	-1.44247
Children	-0.741182
Bias	0.695176
Sex-Мужской	0.269977
Sex-Женский	-0.26996
Class auto-C,D	-0.221445

Рисунок 12 – Вес признаков базы данных¹⁰

¹⁰ Полученные результаты в Microsoft Azure ML

Наиболее значимым признаком является возраст клиента – чем меньше возраст клиента, тем больше вероятность наступления страхового случая. Так же не мало важными признаками наступления страхового случая являются класс автомобиля клиента и наличие детей – чем выше класс автомобиля, тем выше вероятность страхового случая. Так признак, как наличие детей у клиента снижает вероятность наступления страхового случая.

Применив обученную модель на тестируемые данные, получили следующий результат (рис.13).

True Positive	False Negative	Accuracy	Precision
34	27	0.859	0.895
False Positive	True Negative	Recall	F1 Score
4	155	0.557	0.687
Positive Label	Negative Label		
1	0		

Рисунок 13 – Характеристики разработанной модели¹¹

Ассигасу – доля правильных ответов, равна 85,9 %. Precision – мера точности, характеризующая, сколько полученных в результате использования модели классификации положительных ответов являются правильными. В данном случае точность разработанной модели составляет 89,5%, что является высоким показателем [60]. На рисунке 14 представлена ROC – кривая модели.

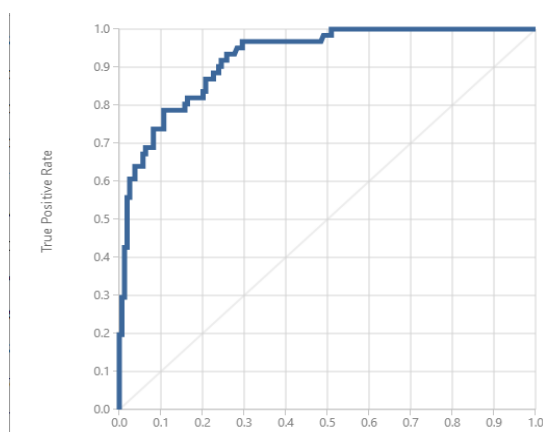


Рисунок 14 – ROC – кривая модели ML¹²

¹¹ Полученные результаты в Microsoft Azure ML

¹² Полученные результаты в Microsoft Azure ML

На рисунке 15 представлены обученные тестируемые данные. «*Scored Probabilities*» - прогнозная вероятность наступления страхового случая. «*Insurance case*» - страховой случай (1 – страховой случай наступил, 0 – страховой случай не наступил).

Sex- Женский	Sex- Мужской	Age	Children	Class auto- A,B	Class auto- C,D	Class auto- E,S	Insurance case	Scored Labels	Scored Probabilities
1	0	51	1	0	1	0	0	0	0.0321
1	0	61	0	0	0	1	0	0	0.343025
0	1	21	1	0	0	1	1	1	0.932536
1	0	52	0	1	0	0	0	0	0.018469
1	0	29	1	0	1	0	1	0	0.183368
0	1	64	1	0	0	1	1	0	0.247519
0	1	29	0	0	0	1	1	1	0.935355
0	1	23	0	0	1	0	1	1	0.576649
0	1	47	1	0	1	0	0	0	0.074566
0	1	62	0	0	0	1	1	0	0.450949
0	1	20	1	0	1	0	0	0	0.457269

Рисунок 15 – Обученные данные прошлых клиентов¹³

Данная таблица позволяет сравнить реальные данные с данными, полученными от обучения моделью. Общий вид разработанной модели машинного обучения представлен на рисунке 16.

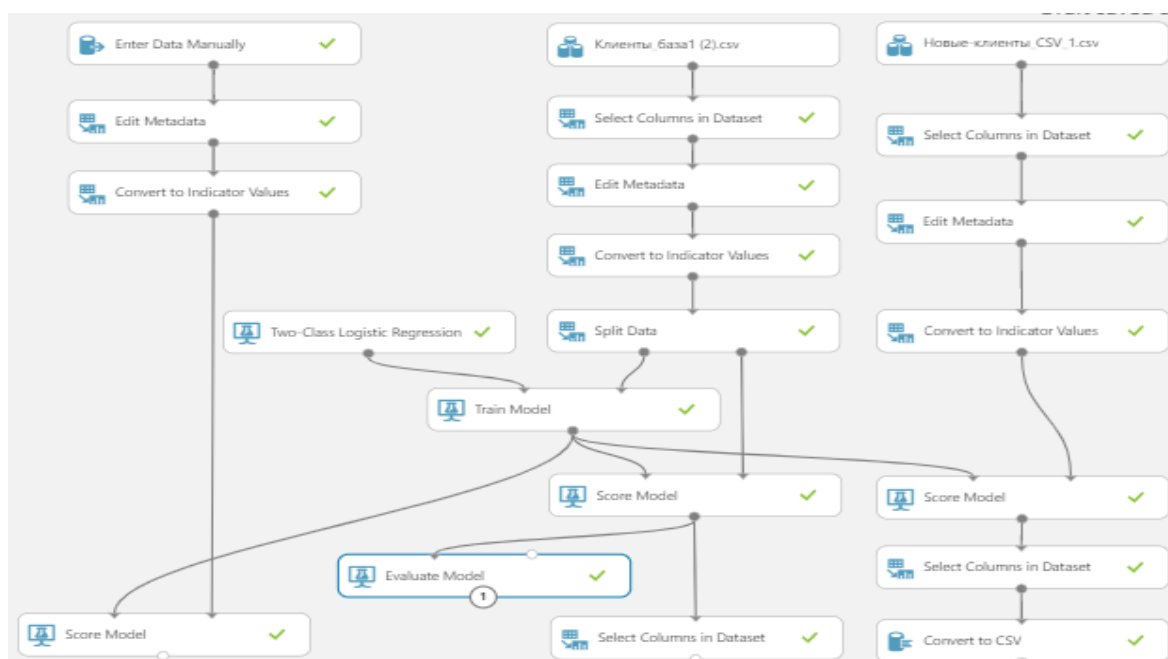


Рисунок 16 – Разработанная модель машинного обучения¹⁴

¹³ Полученные результаты в Microsoft Azure ML

¹⁴ Разработано автором в Microsoft Azure ML

Используя данную модель, можно определить вероятность наступления страхового случая для конкретного нового клиента. Быстрый ввод данных вручную производится с помощью модуля «Enter Data Manually». В данном модуле первая строка означает название столбцов (рис.17).

Data			
1	"Sex",	"Age",	"Children", "Class auto"
2	"Женский",	"45",	"1", "A,B"
3	"Мужской",	30,	"0", "E,S"

Рисунок 17 – Окно ввода данных вручную¹⁵

Например, если страховой полис ОСАГО будет оформлять женщина в возрасте 45 лет, имея детей и автомобиль класса «А, В», то вероятность наступления страхового случая равна меньше 2% (*Scored Probabilities*). Если же полис оформлять будет мужчина в возрасте 30 лет, не имея детей и имея автомобиль класса «Е, S», то его вероятность наступления страхового случая значительно увеличится и будет равна 92% (*Scored Probabilities*) (рис.18).

Sex- Женский	Sex- Мужской	Age	Children	Class auto- A,B	Class auto- C,D	Class auto- E,S	Scored Labels	Scored Probabiliti
1	0	45	1	1	0	0	0	0.016212
0	1	30	0	0	0	1	1	0.929895

Рисунок 18 – Вероятность наступления страхового случая конкретных клиентов¹⁶

Внедрение данной модели ML повысит эффективность использования входных, а также уже имеющихся данных, ускорит процесс обработки

¹⁵ Разработано автором в Microsoft Azure ML

¹⁶ Полученные результаты в Microsoft Azure ML

информационных ресурсов и повысит эффективность деятельности оператора страховой компании в сфере тарификации.

Используя разработанную модель машинного обучения, можно оценить вероятность наступления страхового случая у новых клиентов, тем самым внести изменения в процесс тарификации страховой премии для каждого клиента. Полученные данные импортируются в Microsoft Azure ML для дальнейшей оценке (рис. 19). Количество новых клиентов – 200.

rows 200 columns 7

view as  

	№	Name	Number teaty	Sex	Age	Children	Class auto
							
	1	Farrell	881	Мужской	28	1	A,B
	2	Ridsdale	882	Мужской	28	1	A,B
	3	Farthing	883	Женский	18	0	A,B
	4	Salonen	884	Мужской	61	0	A,B
	5	Hocking	885	Мужской	59	0	C,D
	6	Quick	886	Мужской	62	0	A,B
	7	Toufik	887	Женский	41	1	A,B
	8	Elias	888	Мужской	47	0	C,D
	9	Peter	889	Мужской	42	1	E,S
	10	Cacic	890	Женский	42	1	A,B

Рисунок 19 – Данные о новых клиентах¹⁷

Модель исключает признаки, которые не влияют на наступление страхового случая (*Name*, *Number treaty*). Текстовые данные экспатрированных данных, как и данные прошлых клиентов, преобразовываются в категориальные переменные (рис. 20).

¹⁷ Составлено автором в Microsoft Azure ML

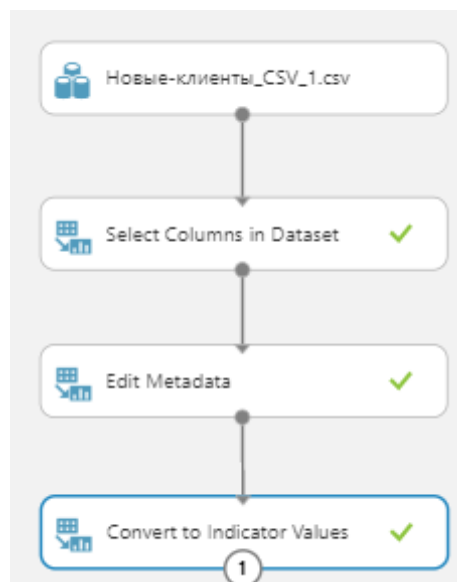


Рисунок 20 – Процесс обработки данных новых клиентов.¹⁸

Полученные данные обученная модель изучает и определяет вероятность наступления страхового случая. Вероятность наступления страхового случая обозначается как «*Scored Probabilities*». (рис. 21).

Sex-Женский	Sex-Мужской	Age	Children	Class auto-A,B	Class auto-C,D	Class auto-E,S	Scored Labels	Scored Probabilities
0	1	28	1	1	0	0	0	0.110284
0	1	28	1	1	0	0	0	0.110284
1	0	18	0	1	0	0	0	0.265563
0	1	61	0	1	0	0	0	0.01455
0	1	59	0	0	1	0	0	0.05622
0	1	62	0	1	0	0	0	0.013355
1	0	41	1	1	0	0	0	0.0228
0	1	47	0	0	1	0	0	0.144624
0	1	42	1	0	0	1	1	0.690119
1	0	42	1	1	0	0	0	0.020941

Рисунок 21 – Вероятности наступления страхового случая у новых клиентов¹⁹

Полученные данные экспортируются из Microsoft Azure ML и используются в расчетах страховой премии для каждого клиента. Принцип расчета страховых премий представлена в таблице 4. Чем выше уровень

¹⁸ Разработано автором в Microsoft Azure ML

¹⁹ Полученные результаты в Microsoft Azure ML

вероятности наступления страхового случая, тем соответственно выше и стоимость страхового полиса ОСАГО для клиента.

Таблица 4 - Уровни вероятности наступления страхового случая²⁰

Уровень вероятности наступления страхового случая	Вероятность страхового случая (Scored Probabilities)	Расчет суммы страховой премии по ОСАГО для нового клиента (Model IP)
<i>Очень низкий уровень</i>	<0,1	Avarage IP + 0 p.
<i>Низкий уровень</i>	0,1 - 0,4	Avarage IP + 500p.
<i>Средний уровень</i>	0,4 - 0,6	Avarage IP + 1000 p.
<i>Высокий уровень</i>	0,6-0,8	Avarage IP + 1500 p.
<i>Очень высокий уровень</i>	0,8-1	Avarage IP + 2000 p.

Avarage IP – средняя страховая премия по ОСАГО. В 2021 году средняя страховая премия по ОСАГО составляет 5 576 Р.

Model IP – сумма страховой премии по ОСАГО для клиента после применения модели машинного обучения. Расчет новых страховых премий представлены в таблицы 55.

Таблица 5- Расчет страховой премии с использованием модели ML²¹

№	Name	Sex	Age	Children	Class auto	Scored Probabilities	Average IP	Model IP
1	Farrell	Мужской	28	1	A, B	0,110	5 576 Р	6 076 Р
2	Ridsdale	Мужской	28	1	A, B	0,110	5 576 Р	6 076 Р
3	Farthing	Женский	18	0	A, B	0,266	5 576 Р	6 076 Р
4	Salonen	Мужской	61	0	A, B	0,015	5 576 Р	5 576 Р
5	Hocking	Мужской	59	0	C, D	0,056	5 576 Р	5 576 Р
6	Quick	Мужской	62	0	A, B	0,013	5 576 Р	5 576 Р
7	Toufik	Женский	41	1	A, B	0,023	5 576 Р	5 576 Р
8	Elias	Мужской	47	0	C, D	0,145	5 576 Р	6 076 Р
9	Peter	Мужской	42	1	E, S	0,690	5 576 Р	7 076 Р
...	...	...	...	...	...	...	...	...
200	Chambers	Женский	55	0	E, S	0,468	5 576 Р	6 576 Р

²⁰ Составлено автором в MS Excel

²¹ Составлено автором в MS Excel

Результаты применения модели машинного обучения в деятельности страховой компании в сфере тарификации представлены в таблице 6.

Таблица 6 - Результат применения модели машинного обучения в сфере тарификации²²

Количество новых договоров ОСАГО -200 шт.	Средняя страховая премия по ОСАГО, руб.	Общая сумма страховых премий по ОСАГО, руб.
До применения модели ML	5 576,00	1 115 200,00
После применения модели ML	6 089,00	1 217 700,00
Разница, руб.	513,00	102 500,00
Разница, %	8,43%	8,42%

Применение разработанной модели увеличило среднюю страховую премию по ОСАГО с 5 576,00 Р до 6 089,00Р, тем самым увеличив общую сумму страховых премий на 102 500,00 Р (8,42 %).

Использование данной модели поспособствует формированию индивидуальной страховой премии для каждого клиента и получению дополнительной прибыли для страховой компании.

При условии внедрения модели, страховой компании будет необходим как минимум один человек, например, главный актуарий, который будет отслеживать и/или корректировать результаты, полученные моделью. Также потребуются дополнительные затраты для настройки технической составляющей модели.

Таким образом, представленный процесс анализа может быть использован страховой компанией в качестве метода оценки вероятности наступления страхового случая на практике.

²² Составлено автором в MS Excel

2.3 Разработка модели полной архитектуры деятельности оператора страхового рынка

Страховая компания «ПАО СК «Росгосстрах» - крупнейшая в России страховая компания, предоставляющая широкий спектр страховых услуг частным лицам и компаниям для защиты от самых разнообразных рисков.

Виды страховых услуг ПАО СК «Росгосстрах»:

- страхование жизни на случай смерти, дожития до определенного возраста или срока либо наступления иного события;
- страхование жизни с условием периодических страховых выплат (ренity) и (или) с участием страхователя в инвестиционном доходе страховщика;
- страхование от несчастных случаев и медицинское страхование;
- страхование средств наземного транспорта, страхование средств железнодорожного транспорта.
- страхование средств железнодорожного транспорта, средств водного транспорта и грузов [61].

Мисси, цели и задачи страховой компании «Росгосстрах» представлены на рисунке 22.



Рисунок 22 – Миссия, цель и задачи компании²³

²³ Составлено автором в MS Visio

Миссия ПАО СК «Росгосстрах» - защита благосостояния людей путем предоставления им доступных и отвечающих их потребностям страховых услуг.

Главная **цель** ПАО СК "Росгосстрах" - стать абсолютным лидером российского рынка страхования, закрепить репутацию надежной, солидной и динамично развивающейся компании.

Стратегическая цель компании — сделать из «Росгосстрах» в течение 10 ближайших лет компанию, которая из года в год будет:

- эталоном надежности и качества обслуживания для клиентов в финансовой отрасли;
- эталоном качества работы сотрудников и качества управления людьми и процессами;
- эталоном работодателя для сотрудников в финансовой отрасли.

Стратегические задачи ПАО СК «Росгосстрах»:

- развитие страховой культуры в России;
- укрепление имиджа надежного, солидного и динамично развивающегося лидера рынка;
- завоевание репутации компании, которой доверяют существующие и потенциальные клиенты.
- предоставление качественных страховых продуктов на всей территории России;
- повышение эффективности системы продаж.

Факторы успеха, стратегические требования и ключевые показатели эффективности страховой компании представлены на рисунке 23.

Бизнес-стратегия ПАО СК «Росгосстрах» заключается в максимальном изучении потребностей клиента, определение спроса со стороны разных клиентских групп и разработка для каждой из них специального предложения.

В настоящее время в компании предлагается довольно широкий набор услуг по автострахованию. Однако, например, в настоящее время не существует предложения, сформированного специально для категории водителей в возрасте до 20 лет, или для владельцев автомобилей красного цвета [62].

И в перспективе страховая компания надеется ответить на самые различные запросы групп своих клиентов во многих линиях бизнеса. Реализовать данный проект по усовершенствованию автострахования поможет внедрение алгоритмов машинного обучения в сфере тарификации [63].



Рисунок 23 - Факторы успеха, стратегические требования и ключевые показатели эффективности страховой компании²⁴

Новая бизнес-стратегия «Росгосстрах» должна стать инструментом, который позволит компании не только укрепить лидерские позиции, но и существенно увеличить долю на российском рынке - с сегодняшних 14% по классическим видам страхования до 25% в ближайшие 5–6 лет.

Эта доля должна быть достигнута за счет развития всех ключевых видов бизнеса, эффективного использования системы многоканальных продаж, а также внедрение IT – инноваций [64].

Ключевыми показателями эффективного развития бизнес-стратегии страховой компании «Росгосстрах» будут являться финансовые показатели, а именно:

²⁴ Составлено автором в MS Visio

- объем страховых премий;
- объем страховых выплат;
- убыточность страховой компании;
- прибыль страховой компании.

Организационная структура ПАО СК «Росгосстрах» представлена на рисунке 24.

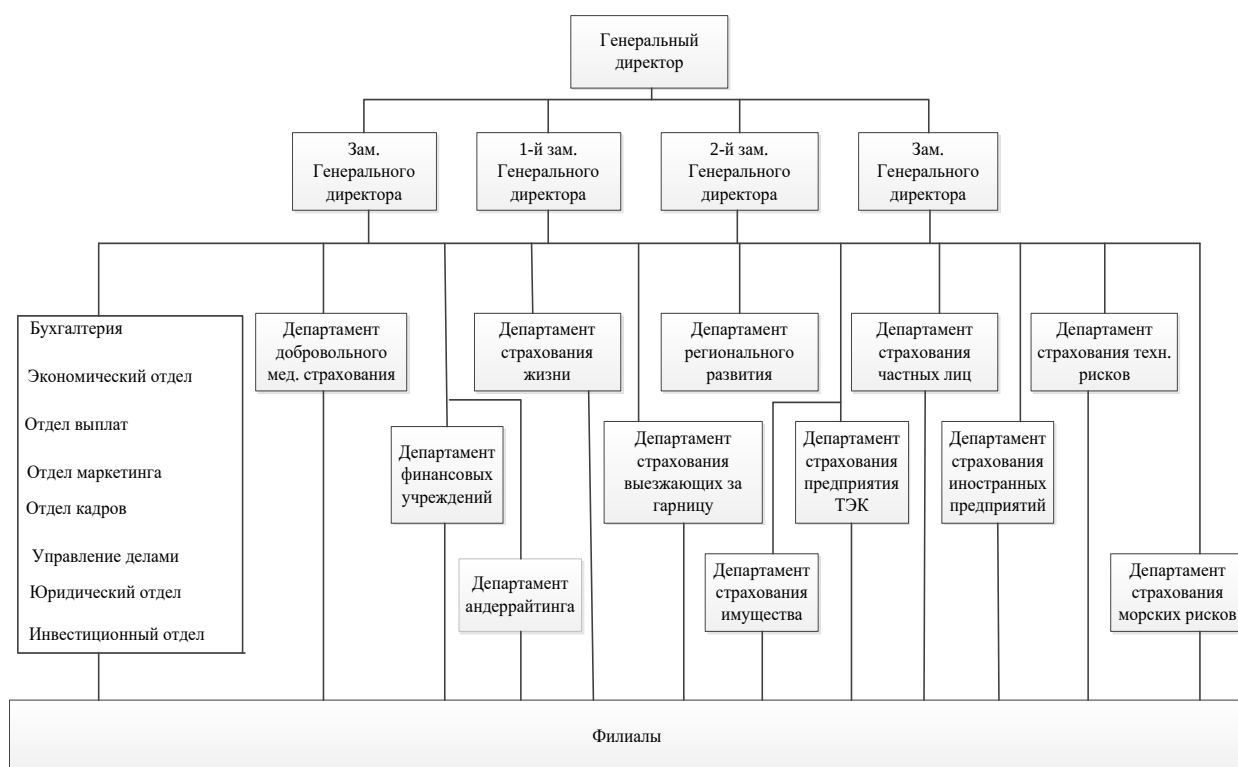


Рисунок 24 – Организационная структура ПАО СК «Росгосстрах»²⁵

Анализируя структуру управления ПАО «Росгосстрах», можно сделать следующие выводы:

- управленческие и производственные процессы взаимодействуют между собой;
- структура центрального аппарата и подразделений компании создается под спроектированные управленческие процессы. В формируемой системе управления важнейшее место занимают знания и навыки владения новыми приёмами управления;

²⁵ Составлено автором в MS Visio

- реализован процессный подход к управлению системы менеджмента качества, отвечающей требованиям МС ИСО 9001:2000 [65].

Для удобства и наглядности представления структуры предприятия построена карта процессов (рис. 25). Деятельность любой страховой компании включает как общие практически для каждой коммерческой организации бизнес-процессы (так называемые вспомогательные), так и специфические для предметной области (в данном случае это страхование) основные бизнес-процессы (функционал деятельности).

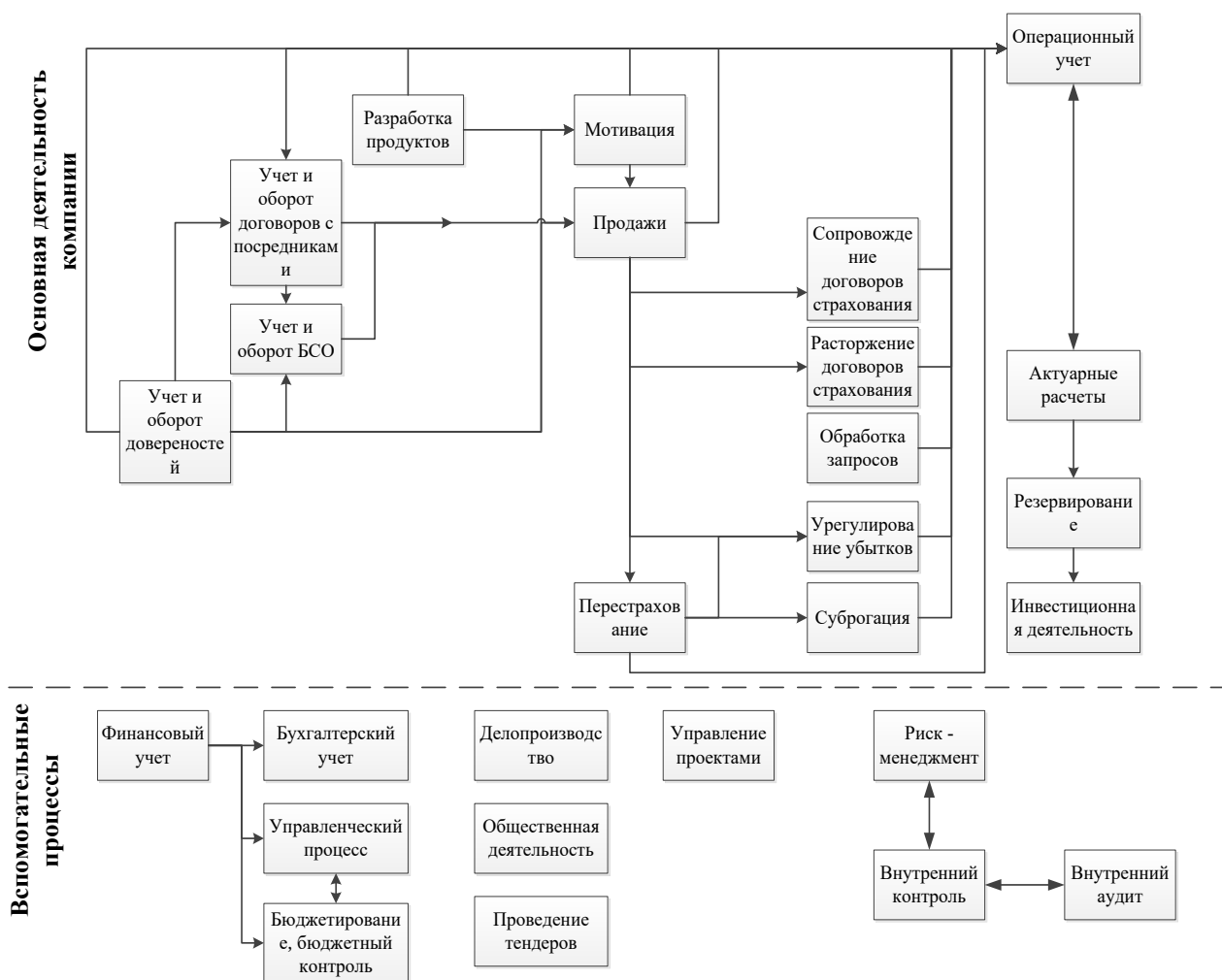


Рисунок 25 - Карта бизнес – процессов страховой компании

«Росгосстрах»²⁶

Важным этапом анализа деятельности организации является выделение и классификация бизнес-процессов. Деятельность страховой фирмы,

²⁶ Составлено автором в MS Visio

декомпозируется на блоки процессов с однородными целями (рис. 26). Бизнес-процессы с однородными по смыслу целями образуют блоки, которые условно можно идентифицировать как:

- 1) блок взаимодействия с клиентами, продажи услуг страхования и урегулирования убытков,
- 2) блок операционного и аналитического характера, обеспечивающий поддержку процессов первой группы,
- 3) блок, обеспечивающий поддержку функционирования страховой фирмы,
- 4) блок процессов управления страховой фирмой в целом [66].

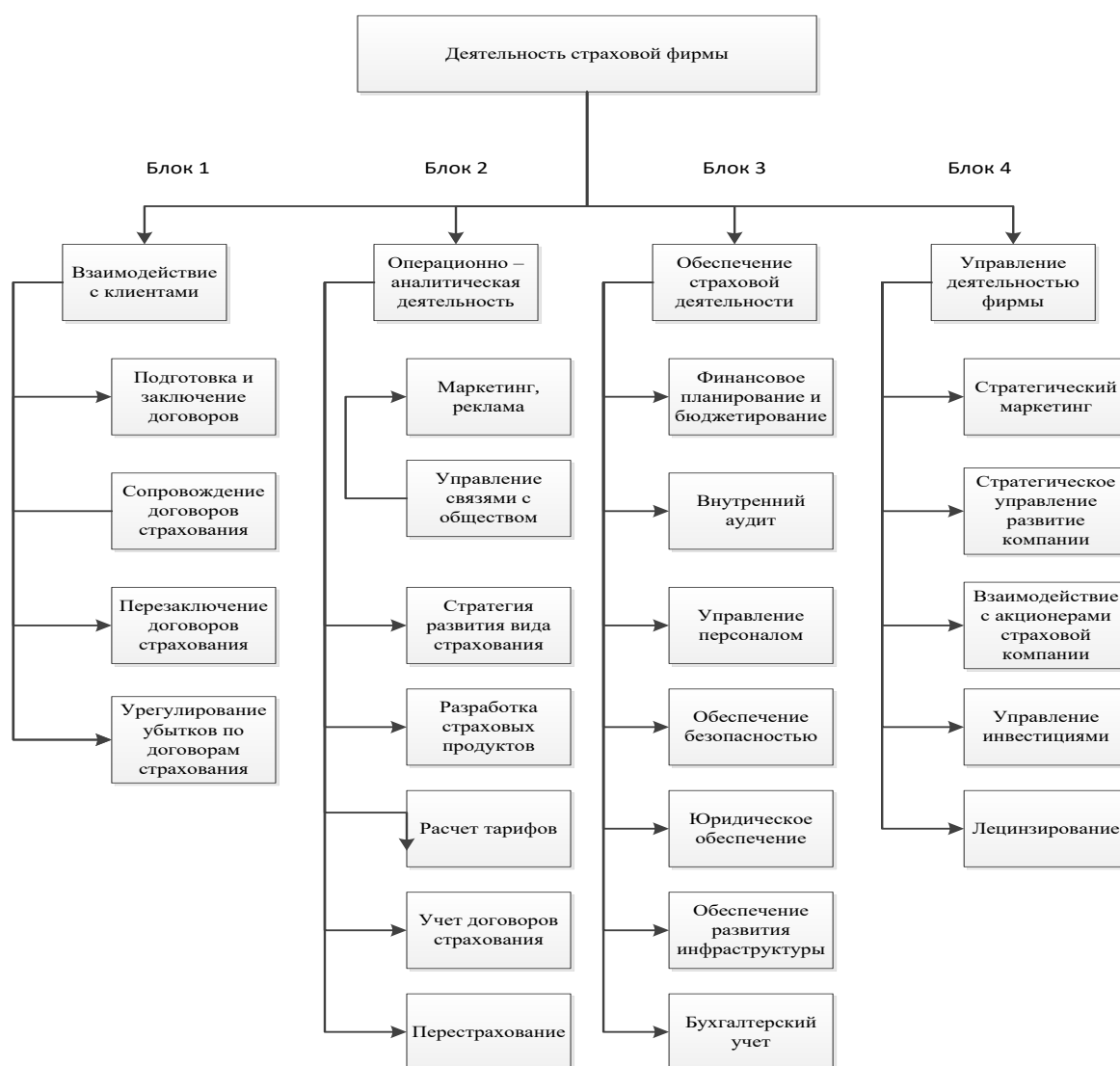


Рисунок 26 – Классификация бизнес – процессов страховой компании²⁷

²⁷ Составлено автором в MS Visio

Для каждого выделенного блока процессов обозначена характерная только для него часть общего внешнего окружения страховой компании.

Модель информационных систем – это совокупность программных продуктов, приложений и сайтов, которые организация использует в своей деятельности [67]. На рисунке 27 изображена модель информационных систем и сервисов страховой компании «Росгосстрах».

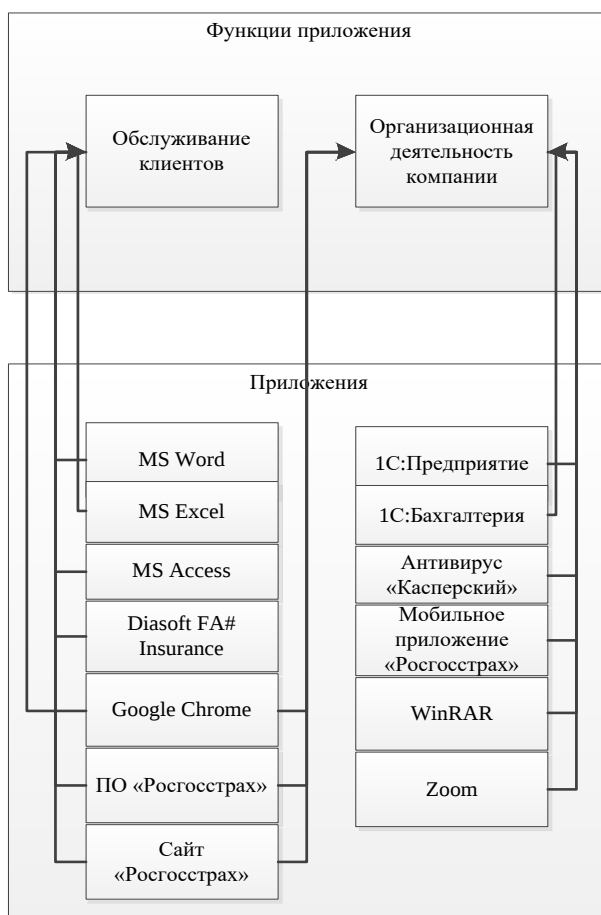


Рисунок 27 – Модель информационных систем «Росгосстрах»²⁸

Организационная деятельность предполагает использование таких программных продуктов, как 1С Предприятие, 1С Бухгалтерия, Антивирус «Касперский», мобильной приложение «Росгосстрах», WinRAR и Zoom для деловых видео встреч.

²⁸ Составлено автором в MS Visio

Для обслуживания клиентов страховая организация использует программное обеспечение «Росгосстрах», Google, MS Excel, MS Word, MS Access, Diasoft FA# Insurance, сайт «Росгосстрах».

На рисунке 28 представлена модель связи приложений и бизнес-процессов страховой компании.

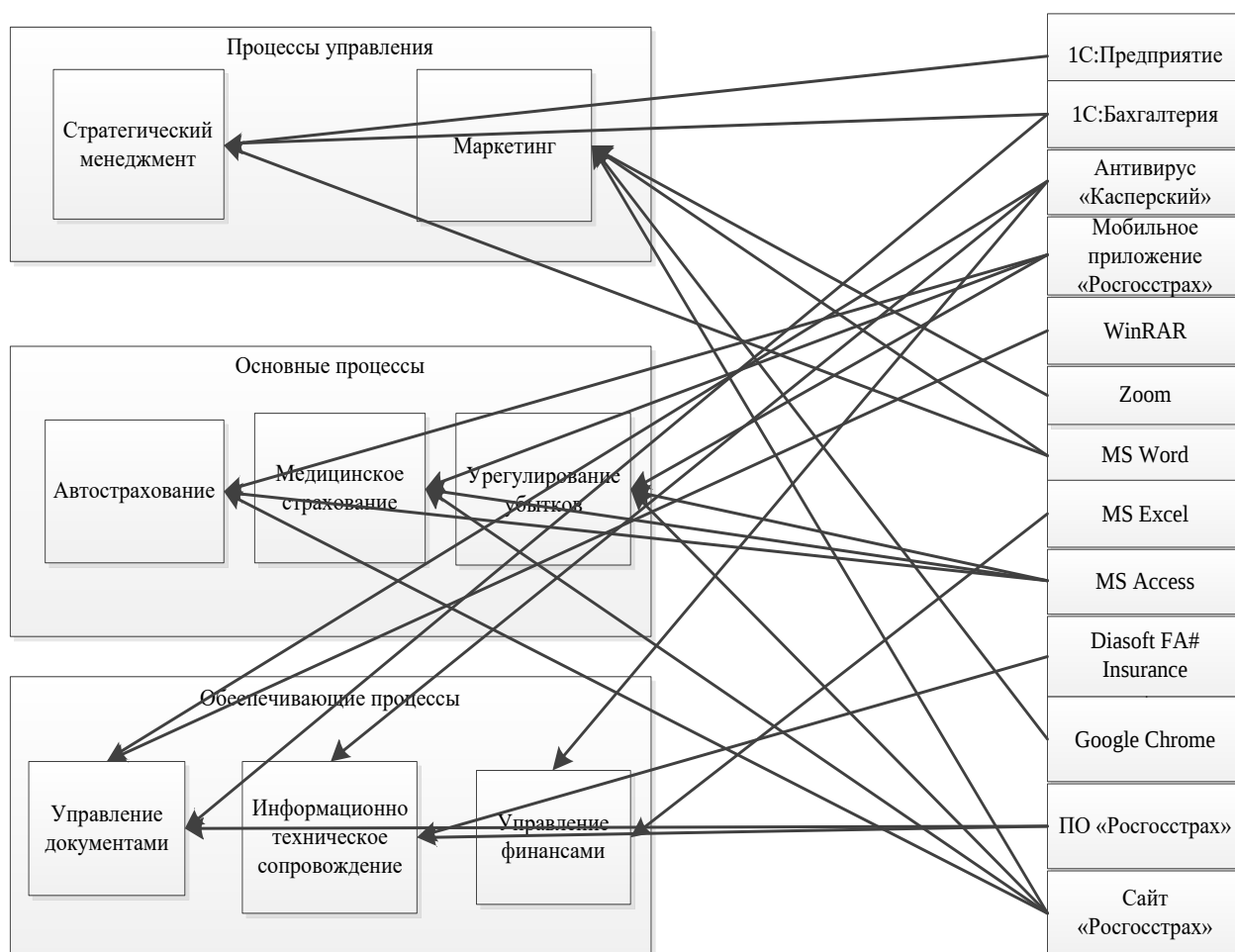


Рисунок 28 – Модель связи приложений и бизнес-процессов страховой компании²⁹

Таким образом, построенная модель связи процессов и приложений демонстрирует, какие приложения задействованы основными процессами организации, процессами управления и обеспечивающими процессами.

На рисунке 29 представлена ИТ – инфраструктура ПАО СК «Росгосстрах».

²⁹ Составлено автором в MS Visio

ИТ-инфраструктура страховой компании представляет собой совокупность технических и технологических решений для обеспечения эффективного функционирования бизнес-процессов предприятия в соответствии с правилами и концепциями, определяемыми бизнес-архитектурой [68].

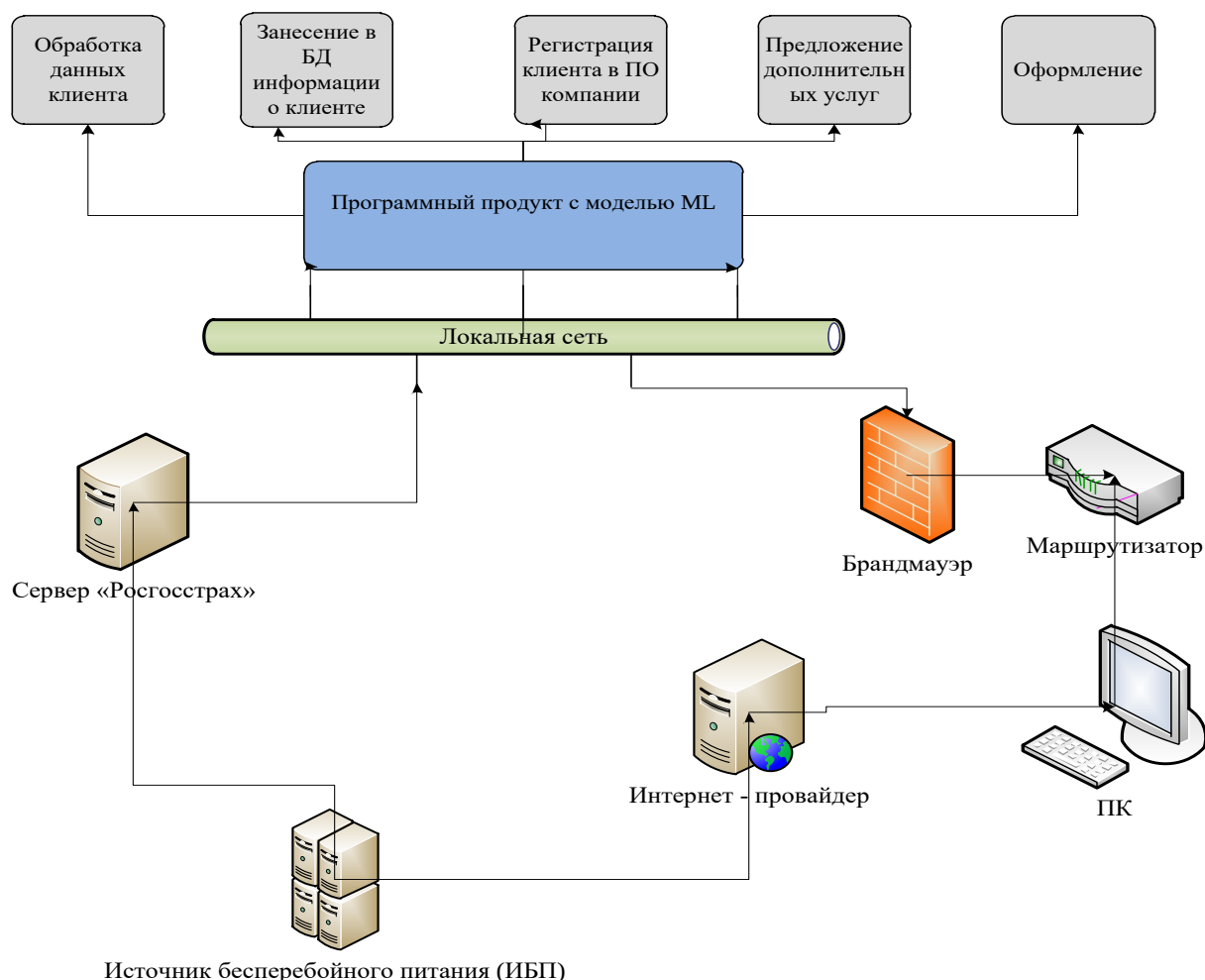


Рисунок 29 – ИТ – инфраструктура ПАО СК «Росгосстрах»³⁰

Полная модель архитектуры организации является важным элементом для синхронизации задач, ресурсов и организационных практик при планировании внедрения моделей машинного обучения в деятельность страховой компании (Приложение А). Таким образом, архитектура ПАО СК «Росгосстрах» описана на всех уровнях. Это обеспечивает целостный и всеобъемлющий взгляд на все аспекты деятельности страховой компании.

³⁰ Составлено автором в MS Visio

2. 4 Результаты и выводы

Наибольшую востребованность у страховщиков показывают технологии «интернетизации» бизнес-процессов» и технологии больших баз данных. Однако при реализации проекта информатизации компании и использовании новых технологий, страховые компании могут столкнуться с недостатками готовых ИТ-решений и отсутствие четкого понимания выгоды от внедрения новых технологий. Основным же барьером на пути развития новых технологий страховщики считают ограничивающую регуляторную среду. И, напротив, недостаток квалифицированных кадров перестал быть серьезным препятствием для менеджмента.

Используя данные о прошлых клиентах страховой компании, была разработана модель машинного обучения. В разработке модели была использована онлайн-платформа Microsoft Azure ML. В основе модели лежит метод логистической регрессии для бинарных переменных. Разработанный алгоритм предсказывает вероятность наступления страхового случая у конкретного клиента, подгоняя данные к логистической функции. Мера точности, характеризующая, сколько полученных в результате использования модели классификации положительных ответов являются правильными, в разработанной модели ML составляет 89,5%, что является высоким показателем.

Описана и разработана полная модель архитектуры предприятия ПАО СК «Росгосстрах» на всех уровнях: бизнес-стратегия, бизнес-архитектура, архитектура приложений, ИТ-инфраструктура. Это позволило рассмотреть все протекающие внутри предприятия бизнес-процессы целиком, а также создать цепочку, показывающую воздействие отдельных элементов стратегии развития предприятия на его бизнес-процессы, и их зависимость от информационных систем и технологических элементов.

3 ОЦЕНКА РЕЗУЛЬТАТИВНОСТИ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ В СТРАХОВОЙ ДЕЯТЕЛЬНОСТИ

3.1 Информационное обеспечение и поддержка технологии машинного обучения в деятельности оператора рынка страховых услуг

Управлять и внедрять проект машинного обучения сложно, поскольку проекты часто носят исследовательский характер, и трудно предсказать, сколько времени потребуется на их завершение. Часто всё начинается с одной идеи, а затем перетекает в новое направление, когда предложенный метод не срабатывает или если предположения относительно данных, оказываются неверными [69].

Проект по внедрению машинного обучения можно разбить на три основных этапа: сбор данных, моделирование и развёртывание. Все они влияют друг на друга. Под моделированием понимается использование алгоритма машинного обучения для поиска информации в собранных данных. Развёртывание берёт набор инструкций и использует его в приложении. Это приложение может быть сервисом для страховой компании, которое пытается наиболее точно определить вероятность страхового случая у конкретного клиента.

Сам же проект по внедрение методов машинного обучения состоит из 7 этапов [70].

1) Оформление документов на согласие о внедрении машинного обучения в деятельность страховой компании. Это начальный этап проекта. Оно включает в себя переговоры со стейкхолдерами, чтобы понять цели и ожидания от проекта, переговоры с бизнес-аналитиками, которые помогают выяснить, какие данные нам доступны и где их можно получить. Улучшить интуитивное понимание задачи поможет формулирование некоторых первоначальных запросов и исследование данных.

2) Сбор информации и формирование базы данных. Это традиционный этап исследования данных с помощью Pandas и Jupiter Notebook с целью

получения каких-либо выводов о используемых данных. Типичный анализ включает подсчет количества строк в данных, создание гистограмм для различных агрегатов функций, графиков тенденций с динамикой и нескольких графиков распределения. Исследователи также будут формировать запросы, которые станут ядром их ML модели.

3) Выбор и разработка алгоритма машинного обучения. В этом собственно вся суть проекта. На этом этапе исследователи начинают построение своих моделей, используя наш внутренний фреймворк. Это включает в себя создание ML-системы, конструирование признаков и подготовку модели. Этап также включает построение базовых моделей и всестороннюю оценку окончательного решения.

4) Тестирование разработанных алгоритмов. На данном этапе производится тестирование разработанного алгоритма и MS – модели.

5) Разработка программного продукта. На данном этапе происходит реализация кода окончательной версии программного продукта. Распространенные задачи на этом этапе включают в себя добавление комментариев ко всем функциям и обеспечение правильного форматирования кода в соответствии со стандартами Python и соглашением рабочей команды.

Код снабжается такими отчетными метриками, как количество извлеченных строк, количество строк в выходных данных, погрешность прогнозирования по нескольким метрикам и важность признаков, когда это возможно.

6) Тестирование и внесение поправок в ПП. Большинство моделей проходят этап тестирования. С его помощью специалисты и стейкхолдеры определяют детали теста: период проведения тестирования, процент трафика, контроль тестирования, интерпретация результатов и т. д.

7) Внедрение программного продукта с алгоритмом машинного обучения. Предоставление разработанного программного продукта с MS – моделями и внедрении их в деятельность страховой компании [71].

Для анализа рассмотрим более подробно каждый из вышеперечисленных

этапов в виде диаграмм бизнес-процессов, выполненных в нотации BPMN.

На рисунке 30 представлена диаграмма процесса «Оформление документов на согласие о внедрении машинного обучения в деятельность страховой компании».

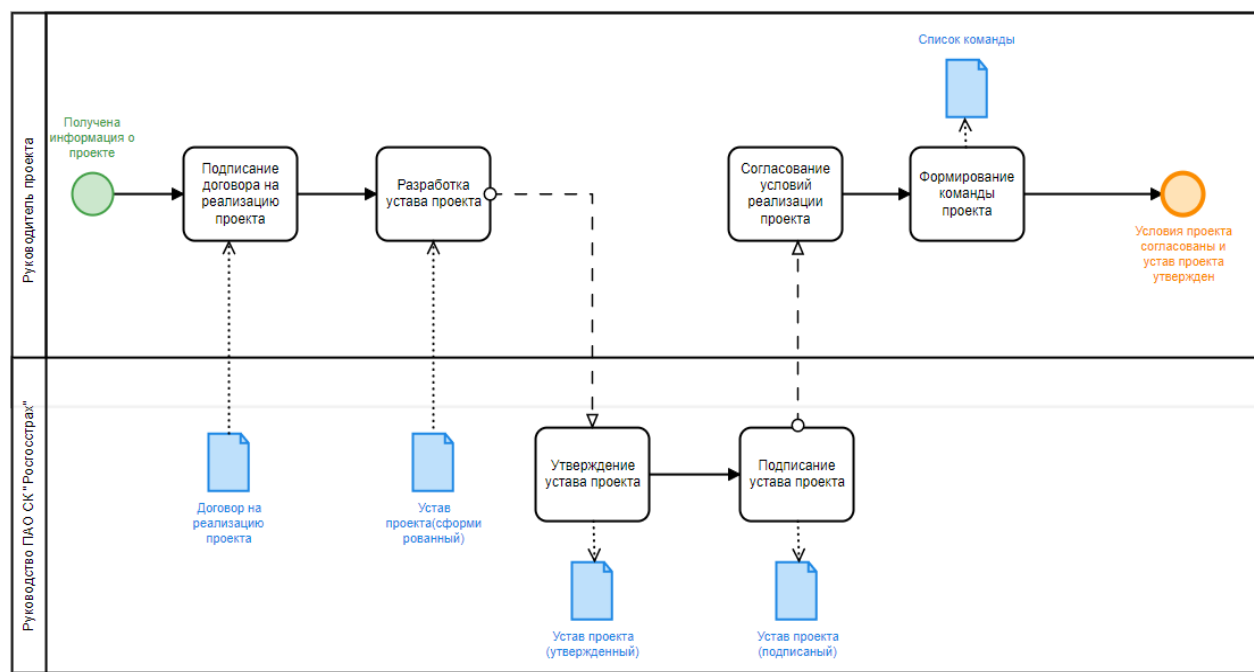


Рисунок 30 – Диаграмма процесса «Оформление документов на согласие о внедрении машинного обучения в деятельность страховой компании»³¹

Началом данного этапа является полученная информация о проекте от заказчика. После чего происходит процесс формирования, утверждения и подписания устава проекта между двумя сторонами – руководством страховой компании и руководителем проекта.

Далее согласуются условия реализации проекта и формируется команда проекта, а также передается информация на следующий этап проекта в виде подписанного устава.

На рисунке 31 представлена диаграмма процесса «Сбор информации и формирование базы данных».

³¹ Составлено автором в нотации BPMN

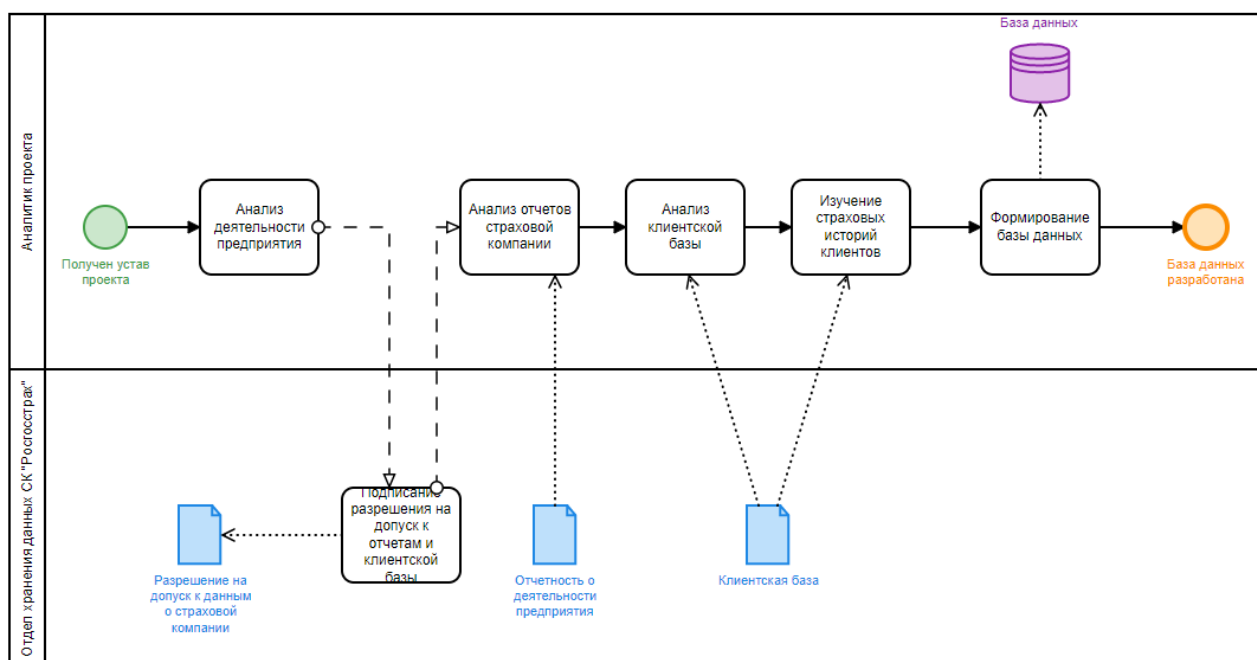


Рисунок 31 – Диаграмма процесса «Сбор информации и формирование базы данных»³²

Из рисунка 31 следует, что после заключения договора на реализацию проекта и подписание устава проводится анализ деятельности страховой компании, а именно:

- анализ отчетностей предприятия;
- анализ клиентской базы;
- анализ страховых историй клиентов.

Далее на базе проведенного анализа следует формирование базы данных для дальнейших этапов реализации проекта.

На рисунке 32 изображена диаграмма процесса «Разработка алгоритма машинного обучения и тестирование». Данный этап является самым важным в процессе реализации проекта.

В данном этапе реализации проекта принимают участие разработчик и тестировщик проекта. После сформированной базы данных аналитиком разработчик приступает к выбору метода машинного обучения. Далее происходит выбор необходимого типа алгоритма и его разработка.

³² Составлено автором в нотации BPMN

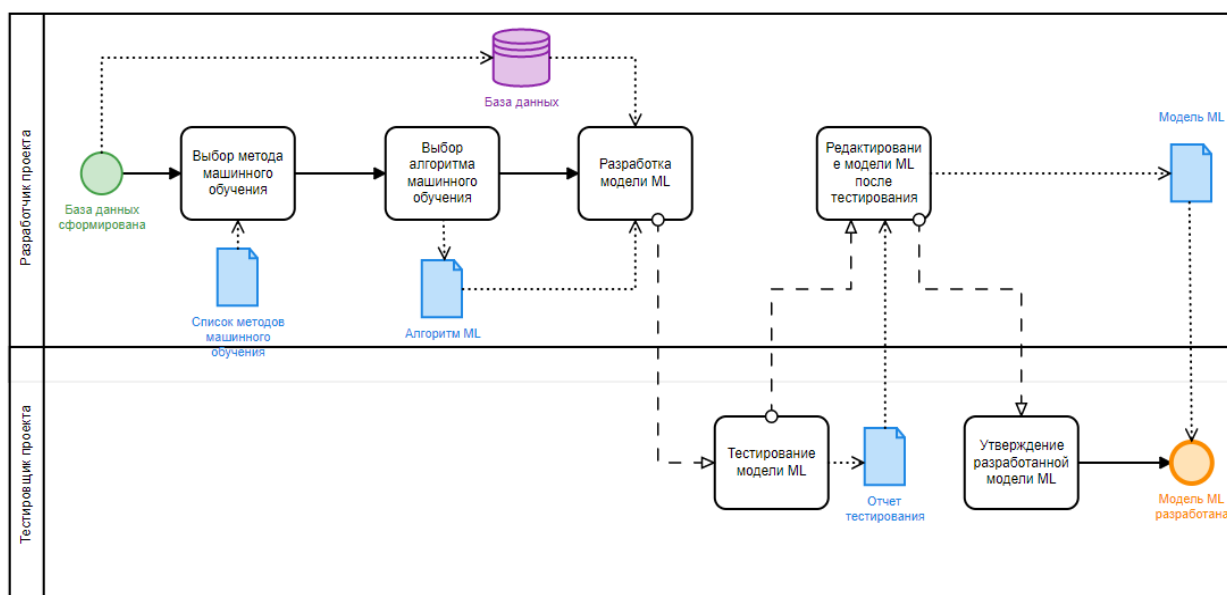


Рисунок 32 – Диаграмма процесса «Разработка алгоритма машинного обучения и тестирование»³³

После этого происходит процесс разработки модели машинного обучения (ML) и ее тестирование тестировщиком. На основании отчета о тестирование разработчик дорабатывает и вносит поправки в модель ML.

В завершении данного этапа происходит утверждение тестировщиком модели и передача необходимых документов на следующий этап реализации проекта.

Следующим этапом реализации проекта по внедрению методов машинного обучения в деятельность страховой компании является процесс разработки программного продукта на основе ранее разработанной ML модели.

На рисунке 33 представлена диаграмма процесса «Разработка программного продукта и тестирование».

Как и в предыдущем этапе, задействованы два специалиста – разработчик и тестировщик проекта. После полученной модели ML, программист – разработчик занимается выбором платформы, на которой будет разрабатываться программный продукт. Далее происходит процесс написания программного кода и выбор стиля оформления программного продукта (персонализация).

³³ Составлено автором в нотации BPMN

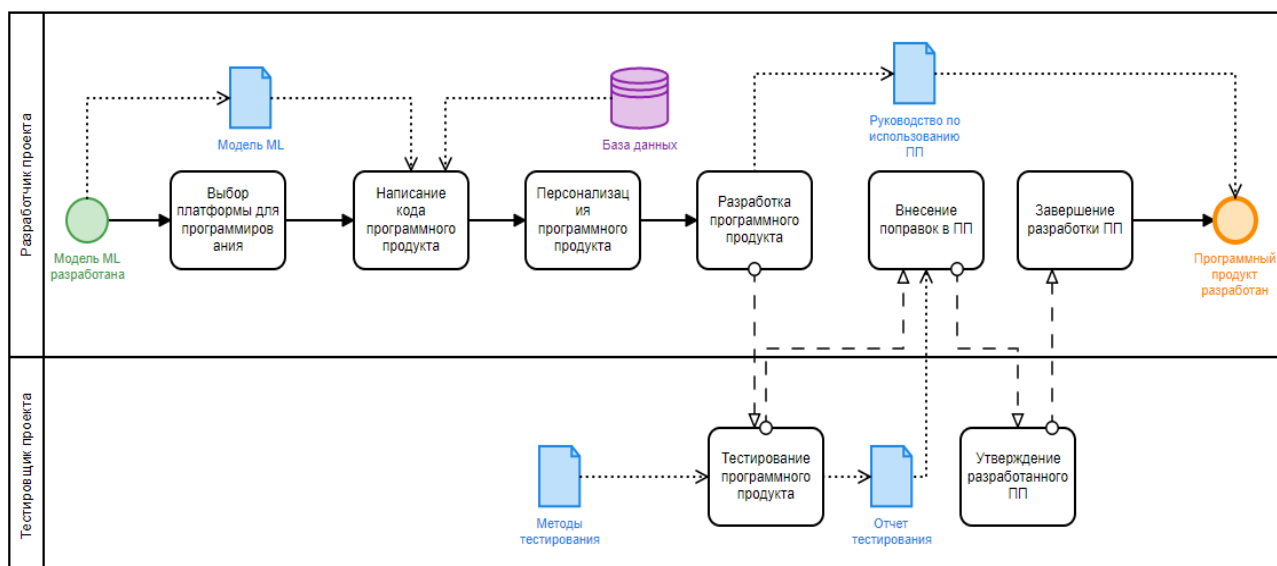


Рисунок 33 – Диаграмма процесса «Разработка программного продукта и тестирование»³⁴

Следующей задачей разработчика является непосредственно сама разработка программного продукта с помощью базы данных и модели ML, которая включает в себя алгоритм машинного обучения. После разработки ПП тестировщик с помощью нескольких методов тестирования производит анализ разработанной программы.

Результатом тестирования является отчёт, который предоставляется разработку для внесения поправок и окончательной доработки программного продукта.

Результатом данного этапа является не только сам программный продукт, но и разработанное программистом руководство по использованию.

Заключительным этапом проекта по внедрению методов машинного обучения в деятельность страховой компании является само внедрение ПП с моделью ML в деятельность страховой компании. На рисунке 34 представлена диаграмма процесса «Внедрение ПП с моделью ML в деятельность страховой компании».

³⁴ Составлено автором в нотации BPMN

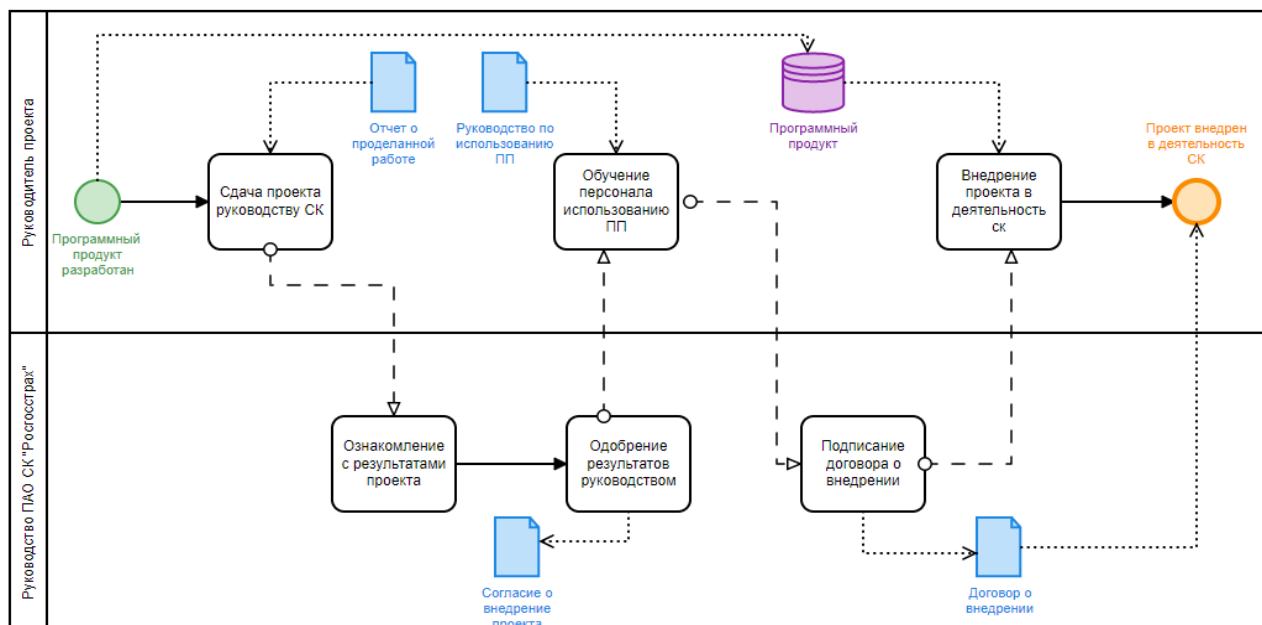


Рисунок 34 – Диаграмма процесса «Внедрение ПП с моделью ML в деятельность страховой компании»³⁵

На последнем этапе руководитель проекта сдает отчет о проделанной работе руководству ПАО СК «Росгосстрах». Руководство компании после ознакомления с отчетом одобряет внедрение проекта и подтверждает это документом о согласии. После чего команда проекта обучает сотрудников компании эксплуатации программного продукта на основании разработанного документа – «Руководство по использованию ПП». После чего подписывается договор о реализации проекта и проект внедряется в детальность страховой компании.

3.2 Оценка влияния алгоритмов машинного обучения в страховании на эффективности работы операторов рынка

Проект внедрения методов машинного обучения в деятельность страховой компании состоит из следующих этапов:

0. Оформление необходимых документов;

³⁵ Составлено автором в нотации BPMN

1. Формирование базы данных;
2. Разработка модели Machine Learning;
3. Тестирование модели Machine Learning;
4. Разработка программного продукта;
5. Тестирование и внесение поправок в программный продукт;
6. Внедрение программного продукта с моделью ML.

На рисунке 35 этапы реализации данного проекта представлены в виде графа.

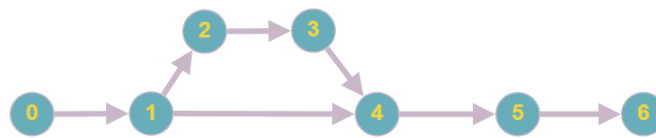


Рисунок 35 – Этапы проекта в виде графа

В разработанной системно-динамической модели уровнями являются невыполненный объём работ по фазе. Выделенные основные уровни и вспомогательные переменные определили вид диаграммы причинно-следственных связей, представленной на рисунке 36.

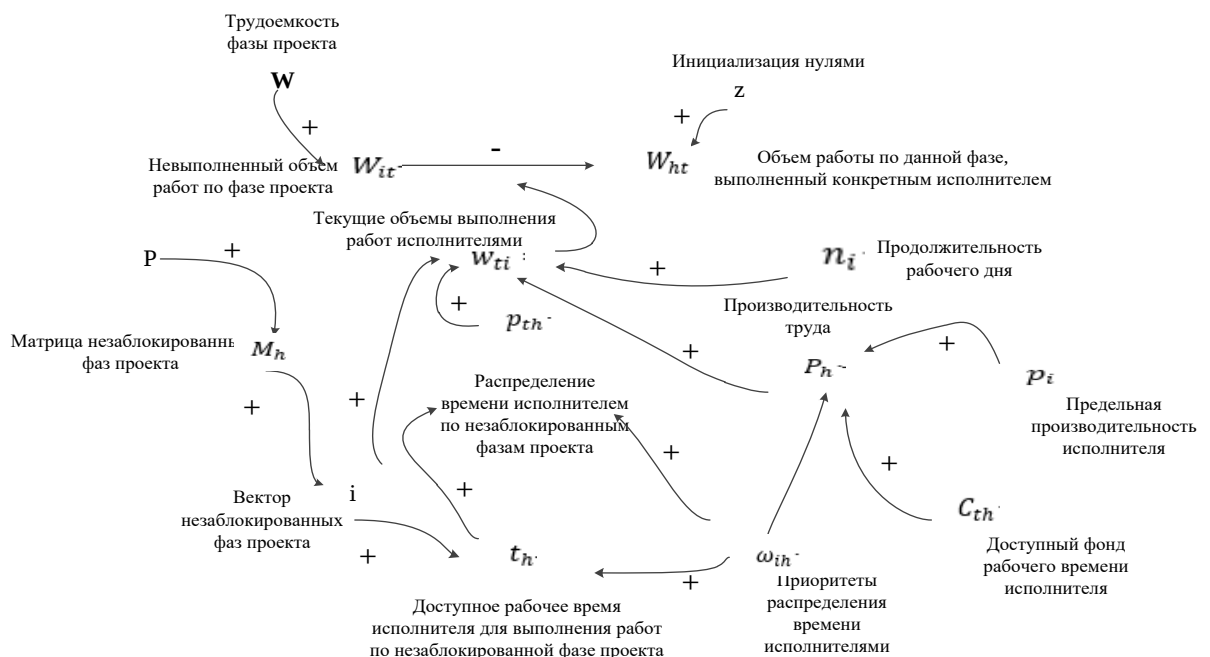


Рисунок 36 – Диаграмма причинно-следственных связей³⁶

³⁶ Составлено автором в MS Visio

В данной модели расчет текущего объема выполнения работ исполнителями производится по формуле:

$$w_{ti} = \sum_{h=1}^M \min(W_{it}, p_{th}, P_h, n_i), i = \overline{1, N},$$

где W_{it} – невыполненный объем работ;

p_{th} – распределение времени исполнителем по незаблокированным фазам проекта;

P_h – производительность труда;

n_i – продолжительность рабочего дня;

i – вектор незаблокированных фаз проекта;

Для определения p_{th} используется формула:

$$p_{th} = 0 \left| C_{th} = 0 \wedge \frac{\omega_{ih}}{\sum_{h=1}^M C_{th}} \right| C_{th} \neq 0, h = \overline{1, M},$$

где C_{th} – Доступное рабочее время исполнителя для выполнения работ по незаблокированной фазе проекта;

ω_{ih} – Доступное рабочее время исполнителя для выполнения работ по незаблокированной фазе проекта;

Расчет n_{tj} производится по формуле:

$$n_{tj} = \prod_{i=1}^N n_{tij}, j = \overline{1, N}$$

При описании проекта необходимо определить объём работ по каждой фазе проекта. В качестве основного ресурса будем рассматривать рабочее время, затрачиваемое каждым сотрудником на выполнение определенной фазы проекта. Для этого таблично зададим (табл.7) приоритеты каждого сотрудника на выполнение определённой фазы проекта с учётом длительности рабочего дня $HRSPerTic$ равной 8 часам. Описание переменных модели представлено в таблице 9.

Таблица 7 - Табличное представление приоритетов сотрудников по фазам проекта³⁷

Worker	Priority								HRS Per week	Productivity			Ex, role
	0	1	2	3	4	5	6				Base	Sigma	
Oliver	0,6						0,4	1	25	1	0,05	0,98457	Менеджер проекта
Jack			0,3	0,1	0,3	0,2	0,1	1	40	1,3	0,05	1,37662	Главный разработчик
Harry		0,1	0,4		0,4		0,1	1	30	1,1	0,05	1,10944	Разработчик
Jacob		0,1	0,4		0,4		0,1	1	30	1,1	0,05	1,09079	Разработчик
Charley		0,8		0,2				1	35	1,2	0,05	1,12027	Аналитик проекта
Thomas	0,2	0,4		0,2			0,2	1	30	1	0,05	1,00695	Бизнес- аналитик
George			0,1	0,1	0,1	0,7		1	30	1,3	0,05	1,23854	Тестировщик проекта
Oscar			0,1	0,1	0,1	0,7		1	30	1,2	0,05	1,15145	Тестировщик проекта
James	0,1	0,1	0,2		0,4	0,1	0,1	1	40	1	0,05	0,97840	Тех. поддержка
William			0,2		0,4	0,2	0,2	1	25	1	0,05	1,05882	Веб- дизайнер
Total									315				

Таблица 8 - Таблица перехода³⁸

Символьное обозначение	Обозначение в Powersim	Сущность переменной
W_{it}	Wrk2Do	Невыполненный объем работ по фазе проекта
W_{ht}	PersWrkPhase	Объем работы по данной фазе, выполненный конкретным членом команды
M_h	No Blockwork	Матрица незаблокированных фаз проекта
w_{ti}	WorkBy	Текущие объемы выполнения работ членом команды
i	NBW	Вектор незаблокированных фаз проекта
p_{th}	Priority	Распределение времени члена команды по незаблокированным фазам проекта
t_h	CurrPrior	Доступное рабочее время члена команды для выполнения работ по незаблокированной фазе проекта

На рисунке 37 изображена модель управления проектом по внедрению методов машинного обучения в деятельность страховой компании.

³⁷ Составлено автором в MS Excel

³⁸ Составлено автором в MS Excel

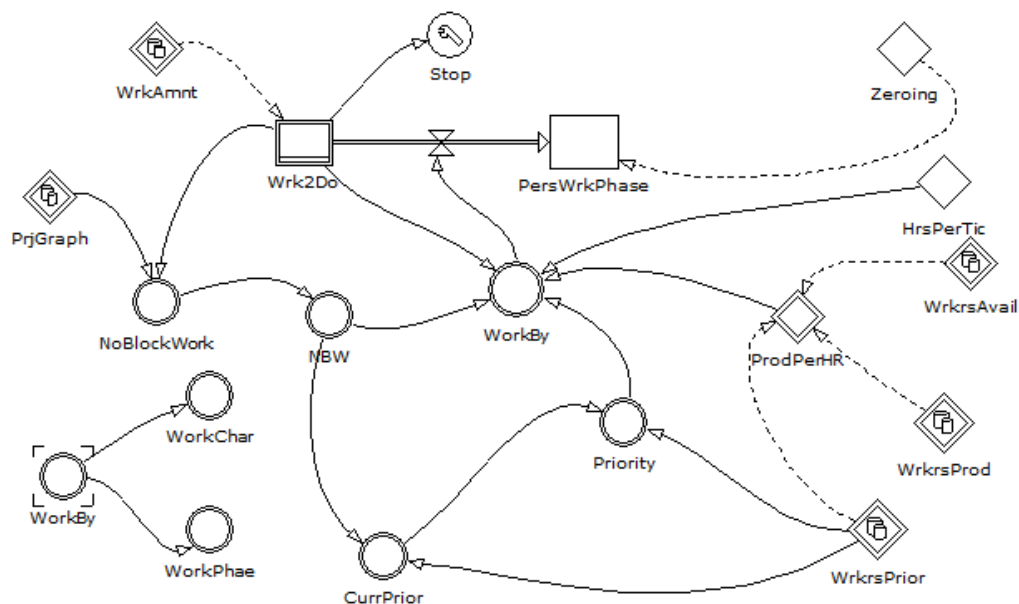


Рисунок 37 – Модель управления проектом внедрения модели ML в
страхование³⁹

Разработанная системно - динамическая модель отражает цикл управления этапами проекта по внедрению методов машинного обучения в деятельность страховой компании. Ключевыми факторами в данной модели являются объём работ по каждой фазе проекта, последовательность фаз проекта, стандартные ставки для персонала проекта, что позволяет определить сроки реализации проекта.

График выполненных работ по фазам проекта представлен на рисунке 38.

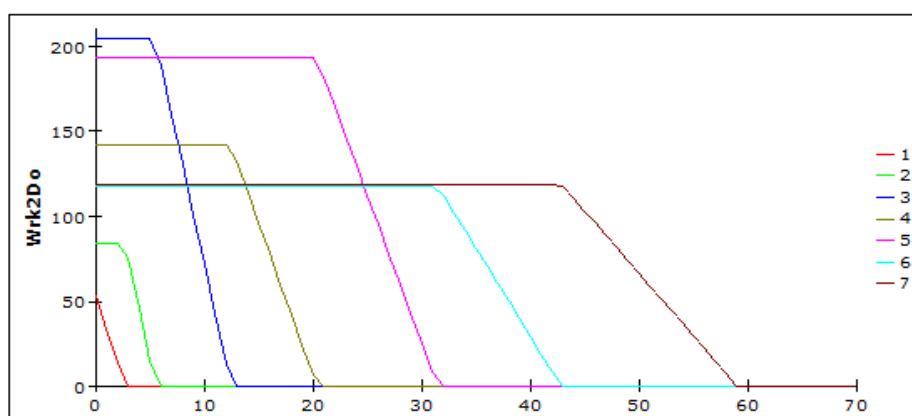


Рисунок 38 – График выполнения работ по проекту⁴⁰

³⁹ Разработано автором в PowerSim Studio 7 Express

⁴⁰ Полученные результаты автора в PowerSim Studio 7 Express

С помощью этого графика можно увидеть, какие работы осуществляются одновременно, а какие последовательно, оценить объем выполненных и невыполненных работ на конкретный период времени.

Анализируя график уровня *Wrk2Do*, представленный на рисунке 38, можно определить, что окончание последней фазы проекта приходится на 58 шаг модели, что равняется 58 дням реализации проекта. График занятости по каждой фазе проекта изображён на рисунке 39.

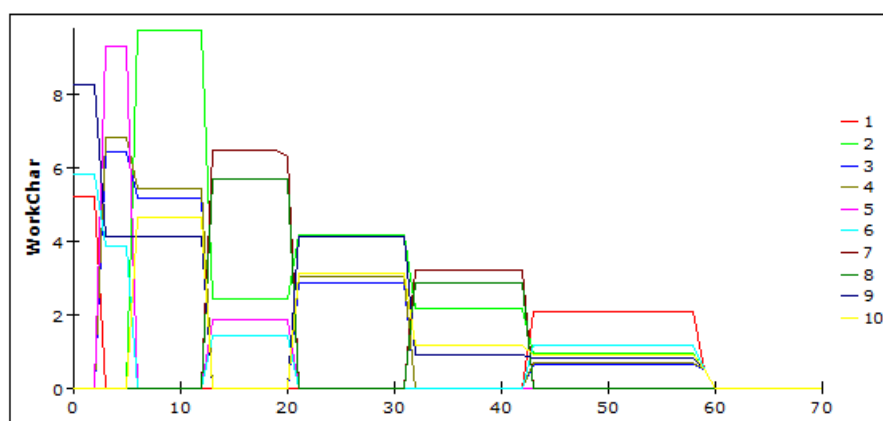


Рисунок 39 – График занятости персонала по каждой фазе проекта ⁴¹

На данном графике показано участие сотрудников и их персональный вклад в исполнение проекта.

На рисунке 40 представлен график накопления трудозатрат персонала по каждой фазе проекта.

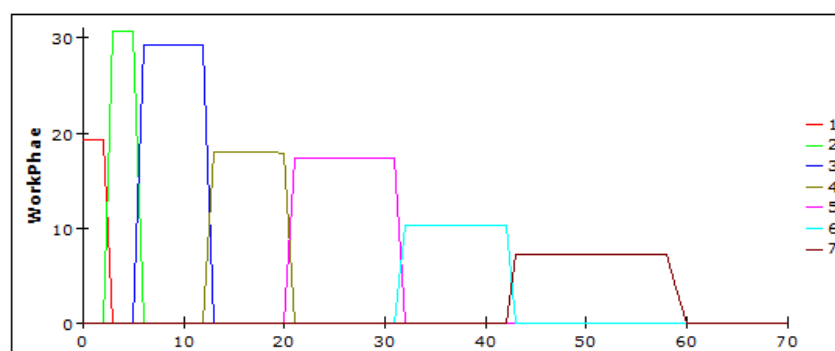


Рисунок 40 - График выполненных работ по фазам проекта⁴²

⁴¹ Полученные результаты автора в PowerSim Studio 7 Express

⁴² Полученные результаты автора в PowerSim Studio 7 Express

Использование разработанной модели позволяет получать управленческие решения с учетом изменяющихся величин входных параметров системы, что обеспечивает непрерывное отслеживание и корректировку эффективности проекта в реальном масштабе времени.

3.3 Оценка экономического эффекта и рисков внедрения технологий машинного обучения в страховании

Команда проекта – совокупность отдельных лиц (участников проекта), привлеченных к выполнению работ проекта и ответственных перед руководителем проекта за их выполнение.

Состав команды включает следующих участников:

1. Руководитель проекта – менеджер, с хорошей технической экспертизой и навыками программирования. Хорошо разбирается в поставленных целях, понимает бизнес-задачу заказчика (страховой компании);

2. Аналитик проекта – проводит анализ, занимается разработкой и формированием необходимой базы данных. Обеспечивает двустороннюю взаимосвязь между предметными экспертами (функциональными специалистами) Заказчика и IT-специалистами путем сбора требований, их обработки, документирования и передачи специалистам, а также путем доведения полученных результатов до представителей страховой компании.

3. Ведущий разработчик проекта. Участвует в проработке и разработке необходимых алгоритмов, моделей и непосредственно самой архитектуры ПП. Также обеспечивает настройку средств визуализации данных (отчетов, запросов, графиков) в необходимом и наиболее эффективном для пользователя виде.

4. Тестировщик проекта. Занимается тестированием все возможных разработок команды проекта и формированием отчета тестирования для дальнейшего внесения поправок разработчиком.

5. Заказчик. Заказчиком в данном проекте является руководство страховой компании «Росгосстрах»

Календарный план проекта представлен на рисунке 41 с помощью Project Expert.

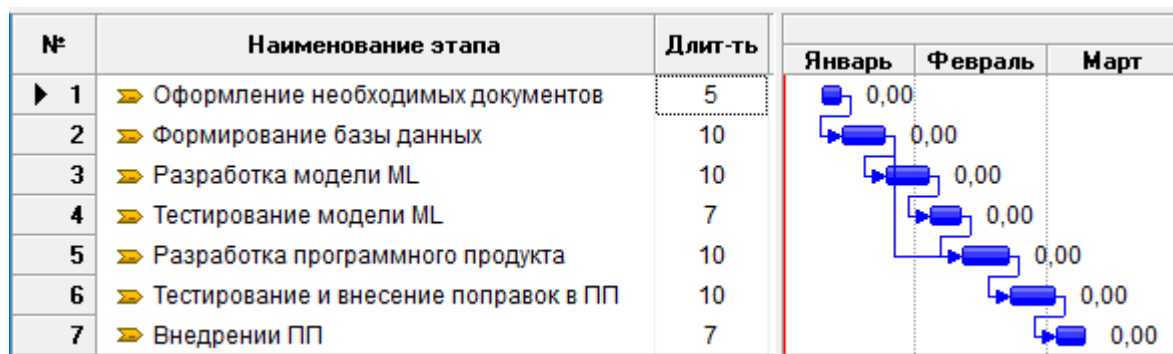


Рисунок 41 - Календарный план реализации проекта⁴³

Этапы и ответственные за этап реализации проекта по внедрению методов машинного обучения в деятельность страховой компании представлены на рисунке 42.

№	Наименование этапа	Длит-ть	Дата начала	Дата окончания	Ответственный
1	Оформление необходимых документов	5	10.01.2022	14.01.2022	Руководитель проекта и заказчик
2	Формирование базы данных	10	15.01.2022	24.01.2022	Аналитик проекта
3	Разработка модели ML	10	25.01.2022	03.02.2022	Разработчик проекта
4	Тестирование модели ML	7	04.02.2022	10.02.2022	Тестирующий и разработчик проекта
5	Разработка программного продукта	10	11.02.2022	20.02.2022	Разработчик проекта
6	Тестирование и внесение поправок в ПП	10	21.02.2022	02.03.2022	Разработчик и тестирующий проекта
7	Внедрении ПП	7	03.03.2022	09.03.2022	Руководитель проекта и заказчик

Рисунок 42 – Ответственные за этапы проекта⁴⁴

Рассмотрим каждый этап более подробно. Этап формирования необходимых документов (5 дней) включает в себя:

- подписание договора на реализацию проекта (1 день);
- формирование устава проекта (2 дня);
- согласование условий проекта (1 день);
- подписание устава проекта и начало реализации (1 день).

Этап формирование базы данных (10 дней) включает в себя:

- анализ деятельности предприятия (2 дня);

⁴³ Составлено автором в Project Expert

⁴⁴ Составлено автором в Project Expert

- анализ отчётов страховой компании (2 дня);
- анализ клиентской базы (2 дня);
- формирование базы данных (4 дня).

Этап разработки модели ML (10 дней) включает:

- выбор метода машинного обучения (2 дня);
- разработка алгоритма (2 дня);
- разработка модели ML (5 дней);
- редактирование модели ML (1 день).

Этап тестирования модели ML (7 дней) состоит из следующих подзадач:

- тестирование модели (4 дня)
- формирование отчета тестирования (1 день)
- утверждение внесенных поправок (2 дня).

Пятый этап – разработка программного продукта (10), начинается только после завершения третьего и четвертого этапов. Данный этап включает в себя:

- выбор платформы программирования (1 день);
- написание кода (4 дня)
- персонализация (1 день);
- разработка программного продукта (4 дня);

Следующий этап – тестирование и внесение поправок в ПП (10 дней) включает:

- тестирование ПП (4 дня);
- формирование отчета тестирования (2 дня);
- внесение поправок (3 дня);
- утверждение доработок ПП (1 день).

Завершающий этап – внедрение ПП (7 дней):

- сдача проекта заказчику (1 день)
- ознакомление с отчётом о проделанной работе (1 день)
- обучение сотрудников компании использованию ПП (2 дня)
- внедрение проекта (3 дня).

Начало проекта приходится на 10.01.2022, а конец на 09.03.2022. Общий

срок проекта составил 59 дней.

Показатели эффективности инвестиционного проекта позволяют определить эффективность проекта по внедрению методов машинного обучения в деятельность страховой компании.

При анализе эффективности инвестиционных проектов используются следующие показатели эффективности инвестиций:

- 1) Чистая приведенная стоимость, NPV;
- 3) Период окупаемости, РВ;
- 3) Внутренняя норма доходности (рентабельности).

В данном случае, при оценке эффективности реализации проекта по внедрению методов машинного обучения в деятельность страховой компании целесообразно будет использовать такой показатель, как чистая приведенная стоимость (NPV).

Чистая приведённая стоимость — это сумма дисконтированных значений потока платежей, приведённых к сегодняшнему дню. Показатель NPV представляет собой разницу между всеми денежными притоками и оттоками, приведёнными к текущему моменту времени.

Формула NPV выглядит следующим образом:

$$NPV = \sum_{t=0}^n \frac{CF_t}{(1 + R)^t}, \text{ где}$$

n, t — количество временных периодов,

CF — денежный поток (Cash Flow),

R — стоимость капитала (ставка дисконтирования, Rate)

Если NPV больше нуля, то проект можно принять, если NPV меньше нуля, то проект стоит отвергнуть.

Выполним расчёт оплаты труда работников по проекту разработки и внедрения моделей ML в деятельность страховой компании (таблица 9).

Таблица 9 - Расходы на оплату труда по проекту⁴⁵

Член команды	Стандартная ставка, руб./час	Количество часов работы, час.	Расход на оплату труда за весь период, руб.
Менеджер проекта	530 руб./час	200 ч.	106 000 руб.
Ведущий разработчик	450 руб./час	320 ч.	126 000 руб.
Разработчик проекта	400 руб./час	240 ч.	96 000 руб.
Аналитик проекта	300 руб./час	280 ч.	84 000 руб.
Специалист по тестированию	350 руб./час	250 ч.	87 500 руб.
Всего			499 500 руб.

Общая стоимость инвестиционного проекта с непредвиденными расходами составляет 550 000 Р.

При условии внедрения модели, страховой компании будет необходим как минимум один человек, например, главный актуарий, который будет отслеживать и/или корректировать результаты, полученные моделью. Рассчитаем эксплуатационные затраты (табл. 10).

Таблица 10 - Эксплуатационные затраты⁴⁶

Член команды	Стандартная ставка, руб./час	Количество часов работы в месяц, час.	Расход на оплату труда за весь период, руб.
Главный актуарий	325 руб./час	200 ч.	65 000руб.
Всего			65 000 руб.

Общие эксплуатационные затраты на оплату труда в месяц составляет 65 000 Р. Ставка дисконтирования (R) равная 10%, заложенная в расчеты приведенной стоимости, отражает допущение о том, что проект характеризуется средним уровнем риска.

После реализации проекта разработки и внедрения модели машинного обучения в деятельность организации, доход страховой компании на конец

⁴⁵ Составлено автором в MS Excel

⁴⁶ Составлено автором в MS Excel

каждого месяца в среднем увеличится на 235 400 Р. Произведём расчет NPV в MS Excel. Результаты расчетов представлены в таблице 11.

Таблица 11 - Результаты расчета показателя NPV ⁴⁷

Месяц	0	1	2	3	4	5	6
Инвестиционные затраты	550 000 Р						
Эксплуатационные затраты		65 000 Р	65 000 Р	68 000 Р	68 000 Р	68 000 Р	68 500 Р
Доход		228 500 Р	232 000 Р	235 750 Р	231 200 Р	238 250 Р	246 800 Р
Чистый результат по проекту	- 550 000 Р	163 500 Р	167 000 Р	167 750 Р	163 200 Р	170 250 Р	178 300 Р
Коэффициент дисконтирования (R = 10%)	1,000	0,909	0,826	0,751	0,683	0,621	0,564
Дисконтированный денежный поток по проекту	-550 000	148 636	138 017	126 033	111 468	105 712	100 646
NPV проекта	180 511						

Так как показатель NPV равен 180 511 Р и больше нуля, то это говорит об успешности проекта. В связи с нынешней внешнеэкономической ситуацией в Российской Федерации рассмотрены ситуации, при которых проект характеризуется повышением уровня риска. Ставка дисконтирования при этом может быть равна 10% и 15%. Результаты расчета NPV при изменении ставки дисконтирования просветлены на рисунке 43.

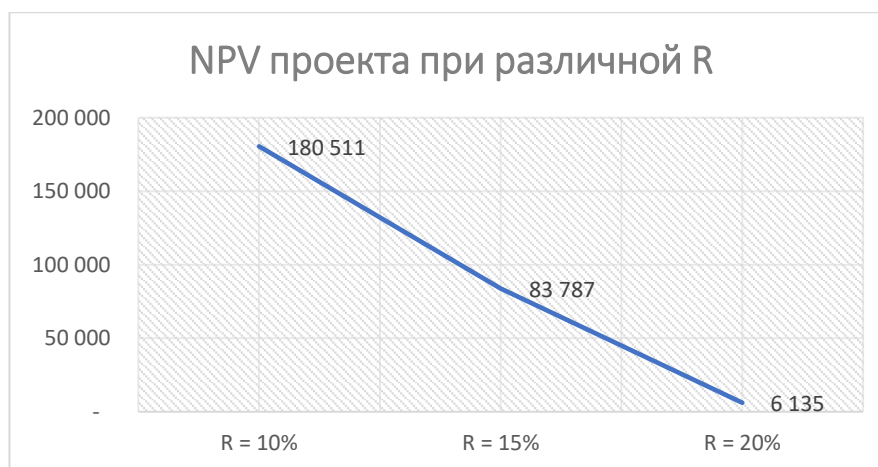


Рисунок 43 – Изменение показателя NPV в зависимости от ставки дисконтирования ⁴⁸

⁴⁷ Составлено автором в MS Excel

⁴⁸ Составлено автором в MS Excel

При $R = 20\%$ NPV проекта равен 6 135 Р и больше единицы, что также говорит об успешности проекта, однако значительно снижает его эффективность.

Срок окупаемости инвестиций в проект был также рассчитан с помощью программы MS Excel. Для этого были использованы дисконтированные денежные потоки из таблицы 100. Из таблицы видно, что дисконтированный денежный поток за 5 месяцев по нарастающей системе составил 629 866 руб., что превышает сумму первоначальных инвестиций. То есть срок окупаемости составит 4 с лишним месяцев. Точный срок окупаемости составит:

$$PB = 4 + \frac{550\,000 - 524\,154}{105\,712} = 4,25 \text{ недель.}$$

Для расчёта индекса доходности проекта спрогнозируем показатель NPV за 12 месяцев и произведем расчёт. Для этого воспользуемся формулой:

$$PI = \frac{\sum_{t=0}^N \frac{CF^t}{(1+i)^t}}{IC} = \frac{592\,865}{550\,000} = 1,08.$$

Таким образом, индекс доходности $PI > 1$. Это означает, что проект заслуживает внимания и подробного анализа на предмет выявления возможных ошибок. Обобщим полученные данные в таблице 12

Таблица 12 - Показатели эффективности инвестиционного проекта

Показатель	Значение
Чистый денежный поток (NPV) при $R = 10\%$, руб.	180 511
Индекс доходности за 12 месяцев	1,08
Срок окупаемости, месяцев	4,25
Ставка дисконтирования, %	10

Таким образом, полученные результаты свидетельствуют, что рассматриваемый проект разработки и внедрения модели машинного обучения в деятельность страховой компании в сфере тарификации является выгодным и может быть принят к реализации.

3. 4 Результаты и выводы

Осуществлен анализ реализации проекта по разработке и внедрению модели машинного обучения в деятельность страховой компании в сфере тарификации с помощью нотации BPMN. В целом проект по внедрению машинного обучения можно разбить на три основных этапа: сбор данных, моделирование и развёртывание. Все эти этапы влияют друг на друга. Также детально рассмотрены участники каждой фазы проекта, а также необходимые для этого документы.

Разработана имитационная модель управления проектом внедрения модели ML в деятельность страховой компании в среде ПП PowerSim Studio 7 Express, позволяющая определить период реализации проекта и объем работ по каждой фазе проекта на основе различных сценариев. Результаты имитационного моделирования показали, что длительность реализации проекта разработки и внедрения модели ML в деятельность страховой компании составит 59 дней.

Осуществлено экономическое обоснование проекта разработки и внедрения модели ML в деятельность страховой компании. Выяснено, что инвестиционные затраты на его реализацию потребуются в размере 550 000 руб. Эксплуатационные затраты в среднем в месяц будут составлять 67 083 руб. Рассчитаны основные показатели эффективности инвестиций: NPV, индекс доходности, срок окупаемости. Полученные результаты анализа свидетельствуют, что рассматриваемый проект является выгодным и может быть принят к реализации.

ЗАКЛЮЧЕНИЕ

В процессе достижения цели исследования решены поставленные задачи и сформулирован ряд выводов.

На основании методов системного анализа изучена технология использования машинного обучения в деятельности страховой компании. Рассмотрены основные направления машинного обучения, которые могут быть использованы в различных сферах страховых услуг. Сфера тарификации выделена как наиболее значимая и учитывая высокий уровень пилотных проектов именно данная группа бизнес-процессов требует внимания аналитиков.

На основании методов системного анализа изучены методы машинного обучения, что позволило систематизировать базовые алгоритмы обучения модели управления данными с учетом специфики бизнес-процессов страховой компании. Сформирована математическая постановка задачи кластеризации в алгоритмах машинного обучения.

На основе анализа научных трудов в сфере машинного обучения определён наиболее подходящий метод машинного обучения в сфере тарификации, что послужило теоретической основой исследования. Предикативное моделирование используется в страховании для определения вероятности наступления страхового случая с тем или иным клиентом на основе страховых историй прошлых клиентов.

На основании методов системного анализа проведен анализ проблем функционирования и развития страхового рынка, что послужило основой для разработки модели машинного обучения.

На основании методов машинного обучения исследования разработан алгоритм модели ML в сфере тарификации страховой компании, которой способен определить вероятность наступления страхового случая конкретно нового клиента на основе имеющихся данных о прошлых клиентах компании. Применении данной модели в сфере тарификации продемонстрирован рост ключевых показателей (средняя сумма страховой премии, общая сумма

страховых премий), что свидетельствует о целесообразности и эффективности реализации проекта. Данная аналитическая модель машинного обучения разработана с помощью веб-платформы Microsoft Azure ML.

На основании архитектурного подхода построена полная модель архитектуры ПАО СК «Росгосстрах», включающая описание уровня стратегии развития, уровня архитектуры процессов, уровня архитектуры приложений и уровня IT-инфраструктуры. Это обеспечивает целостный взгляд на деятельность страховой компании, ее ресурсы и возможности, а также является важным элементом для синхронизации задач, ресурсов и организационных практик при планировании разработки и внедрения модели ML.

На основании процессного подхода детально рассмотрены все этапы проекта разработки и внедрения модели машинного обучения в деятельность страховой компании в нотации BPMN, что позволило определить участников проекта и необходимые документы.

На основании методов математического моделирования осуществлена постановка задачи имитационного моделирования управления проектом разработки и внедрения модели ML в деятельность компании, включающая разработку диаграммы причинно-следственных связей. Осуществлена реализация имитационной модели управления проектом разработки и внедрения модели ML в деятельность компании в среде ПП PowerSim Studio 7 Express, позволяющая определить длительность реализации проекта и объем работ по каждой фазе проекта на основе различных сценариев.

Осуществлено экономическое обоснование внедрения модели ML в сфере тарификации для страховой компании. Полученные результаты свидетельствуют об экономической эффективности предлагаемого проекта.

Дальнейшие перспективы исследования использования модели машинного обучения в сфере страхования связаны с практической апробацией полученных результатов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Цифровая Россия: новая реальность [Электронный ресурс] // McKinsey. – 2018. – URL: <https://www.mckinsey.com/~media/mckinsey/locations/digital-russia-report.ashx> (дата обращения: 25.02.2022).
2. Асадуллина А.В. Цифровая экономика в России: текущий статус и проблемы развития / А.В. Асадуллина // Российский внешнеэкономический вестник. – 2018. – №6. – С. 98-112
3. Страховой рынок России в 2019 году. Национальное рейтинговое агентство [Электронный ресурс]. – URL: https://www.ra-national.ru/sites/default/files/Insurance_Russia_2019_Report_26032020_0.pdf
4. Репин В. В. Процессный подход к управлению. Моделирование бизнес-процессов [Текст] /Репин В.В., Елиферов В.Г.. – 6-е изд.– М.: РИА «Стандарты и качество», 2018.–408 с.
5. Елиферов В.Г. Бизнес–процессы: Регламентация и управление: Учебник [Текст] / Серия учебников для программы МВА.
6. Годин В. В. Управление информационными ресурсами [Текст] / В.В. Годин, И. К. Корнеев – М.: ИНФРА-М, 2015.
7. Гринберг А. С. Информационный менеджмент [Электронный ресурс] / А. С. Гринберг, И. А. Король, – М.: ЮНИТИ-ДАНА, 2017.
8. Костров А. В. Основы информационного менеджмента [Текст] / А.В. Костров. – М.: Финансы и статистика, 2018.
9. Симионов Ю. Ф. Информационный менеджмент [Текст] / Ю.Ф. Симионов, В. В. Бормотов. – Ростов-на-Дону: Феникс, 2016.
10. Хорошилов А. В. Управление информационными ресурсами [Текст] / А. В. Хорошилова, С. Н. Селетков, Н. В. Непровская – М.: Финансы и статистика, 2016.
11. П. Домингос Верховный алгоритм. Как машинное обучение изменит наш мир [Текст] / Домингос П. -М.: Манн, Иванов и Фербер, 2018.-36 с.
12. П. Флах Машинное обучение [Текст]/ -М.: ДМК-Пресс, 2018 г.-40 с.

13. Х. Бринкс, Р. Джозеф, М. Феверол Машинное обучение [Текст] / - М.: Питер, 2019 г.-336 с.

14. Плоткина А. А. Использование методов машинного обучения в имущественном страховании [Электронный ресурс] / А. А. Плоткина // Современная математика и концепции инновационного математического образования. - 2019. - Т. 6, № 1. - С. 260–266. - URL: <https://elibrary.ru/item.asp?id=39241227> (дата обращения: 25.04.2022).

15. Стефанова Н. А. Сущность машинного обучения, его перспективы и развитие [Электронный ресурс] // Н. А. Стефанова, А. С. Трофимец // Актуальные вопросы современной экономики. - 2019. - Т. 1, № 3. - С. 48–52. - URL: <https://elibrary.ru/item.asp?id=39143496>. (дата обращения: 10.05.2022).

16. Использование машинного обучения в ситуационном управлении применительно к задачам электроэнергетики [Электронный ресурс] / Л. В. Массель, О. М. Гергет, А. Г. Массель, Т. Г. Мамедов // Информационные и математические технологии в науке и управлении. - 2019. - № 3 (15). - С. 5-17. – URL: <https://elibrary.ru/item.asp?id=41318513>. (дата обращения: 25.04.2022).

17. Перспективы цифровизации страхового дела в России [Электронный ресурс]. –URL: [05_3_2020_v12.pdf \(finjournal-nifi.ru\)](https://finjournal-nifi.ru/05_3_2020_v12.pdf) (дата обращения: 15.05.2022).

18. Ожидания от машинного обучения в страховании [Электронный ресурс]. – URL: <http://www.asn-news.ru/post/955> (дата обращения: 15.05.2022).

19. Будущее искусственного интеллекта в страховании [Электронный ресурс]. – URL: <https://www.zfort.com.ua/blog/budushee-iskusstvennogo-intellekta-v-strakhovanii> (дата обращения: 25.03.2022).

20. Роль цифровых технологий в трансформации рынка страховых услуг [Электронный ресурс]. – URL: <https://cyberleninka.ru/article/n/rol-tsifrovyyh-tehnologiy-v-transformatsii-rynka-strahovyh-uslug> (дата обращения: 25.04.2022).

21. Умный полис. Как машинное обучение изменит рынок страхования [Электронный ресурс]. – URL: <https://www.forbes.ru/biznes/355131-umnyy-polis->

[kak-mashinnoe-obuchenie-izmenit-rynok-strahovaniya](#) (дата обращения: 08.04.2022).

22. Искусственный интеллект в страховании: чем он может быть полезен [Электронный ресурс]. – URL: <https://www.insur-info.ru/press/159214/> (дата обращения: 17.05.2022).

23. Машинное обучение для страховой компании. Реальные идеи [Электронный ресурс]. –URL: <https://habr.com/ru/company/microsoft/blog/329082/> (дата обращения: 15.05.2022) Гурдус В. Новые технологии и новые ценности [Текст] / В. Гурдус // П., 2018. – с. 1- 13..

24. Директор по цифровым технологиям — chef digital officer (cdo) [Электронный ресурс]. – URL: <http://koptelov.info/publikatsii/cdo-chef-digital-officer/> (дата обращения: 17.05.2022)

25. Цифровизации страхового бизнеса в России [Электронный ресурс]. – URL: <http://www.sluchay.ru/> (дата обращения: 17.05.2022)

26. Ильин И.В. Цифровая трансформация как фактор формирования архитектуры и ИТ- архитектуры предприятия [Текст] / И.В. Ильин, А.И. Левина, А.С. Дубгорн // Научный журнал НИУ ИТМО. – Санкт – Петербург, 2019.- №3. – с. 50 – 55.

27. Кудрявцев Д.В. Архитектура предприятия: переход от проектирования ит - инфраструктуры к трансформации бизнеса [Текст] / Д.В. Кудрявцев, М.Ю. Арзуманян // Журнал менеджмента Russian Management Journal. – Санкт – Петербург, 2017. - № 2. – с. 193 -224.

28. Бабкин Э. А. Анализ непротиворечивости моделей архитектуры предприятия с использованием формальных методов верификации [Текст] / Э.А. Бабкин, Н.О. Понамарев // Информационные системы и технологии в бизнесе. – Нижний Новгород, 2017. –с. 30 – 40.

29. Анализ рисков для страховой компании [Электронный ресурс]. – URL: <https://prana-system.com/vozmozhnosti/dlya-strahovyh-kompanij-i-ih-> (дата обращения: 27.04.2022)

30. Шафрин Ю.А. Основы компьютерной технологии [Электронный ресурс] / А.Ю.Шафрин. – М.: АБФ, 1996. – 202 с. (дата обращения: 21.05.2020).

31. Big Data и Machine Learning против мошенничеств в сфере страхования [Электронный ресурс]. – URL: <https://www.bigdataschool.ru/blog/insurance-cases.html> (дата обращения: 25.04.2022).

32. Щербакова Н.В. Страхование в эпоху цифровых и интернет – технологий [Текст] / Н.В. Щербакова, Ю.М. Ильиных // Экономика. Профессия. Бизнес. – Барнаул. -2019. - № 1, с. 83-86.

33. Котлобовский И.Б. Цифровые технологии и страхование: эволюция отрасли [Текст] / И.Б. Котлобовский // Устойчивое развитие и эволюция страхования в цифровой экономике. – М., 2019. – с. 1-30.

34. Гурдус В. Новые технологии и новые ценности [Текст] / В. Гурдус // П., 2018. – с. 1- 13.

35. Директор по цифровым технологиям — chef digital officer (cdo) [Электронный ресурс]. – URL: <http://koptelov.info/publikatsii/cdo-chef-digital-officer/> (дата обращения: 22.04.2022).

36. Штейнгарт Е.А, Обзор и сравнительная характеристика методологий разработки архитектуры предприятий [Электронный ресурс] / Е.А. Штейнгарт, А.Н. Бурмистров // Экономика и менеджмент предприятия – 2016. – URL: <https://cyberleninka.ru/article/n/obzor-i-sravnitelnaya-harakteristika-metodologiy-razrabotki-arhitektury-predpriyatiy> (дата обращения: 30.04.2022).

37. Сферы применения предикативной аналитики [Электронный ресурс]. – URL: <https://www.ramax.ru/press-center/articles/prediktivnaya-analitika/> (дата обращения: 17.05.2022)

38. Коссов В. В. Методические рекомендации по оценке эффективности инвестиционных проектов (Вторая редакция) [Текст] / В.В. Коссов. - М.: ОАО "НПО "Изд-во "Экономика", 2016. - 421 с.

39. Скрипкин К.Г. Экономическая эффективность информационных систем [Текст] – М.: ДМК Пресс, 2002. – 256 с.

40. Антопольский, А.Б. Информационные ресурсы России: науч. – мет. пособие [Текст] / А.Б. Антопольский. – М.: Либерия, 2004. – 423 с.
41. Ясенев В.Н. Информационные системы и технологии в экономике [Текст] / В.Н. Ясенев. - М.: ЮНИТИ, 2014. - 560 с.
42. Интернет вместо агента [Электронный ресурс]. - URL: <https://rg.ru/2014/07/22/techno.html> (дата обращения: 05.05.2022).
43. Тренды развития ИТ в страховании [Электронный ресурс]. - URL: <http://www.tadviser.ru/index.php/> (дата обращения: 05.05.2022).
44. Хороший клиент — довольный клиент, или зачем Insure нужен Tech [Электронный ресурс]. - URL: <https://www.banki.ru/news/columnists/?id=10902743> (дата обращения: 05.05.2022).
45. Применение методов машинного обучения при оценке размера страховых резервов [Электронный ресурс]. - URL: <https://cyberleninka.ru/article/n/primenenie-metodov-mashinnogo-obucheniya-pri-otsenke-razmera-strahovyh-rezervov> (дата обращения: 18.04.2022).
46. Разработка моделей ML в страховании [Электронный ресурс]. - URL: <https://www.castcom.ru/services/web-development> (дата обращения: 18.04.2022)
47. Цифровая «жизнь» - как новые технологии меняют страхование жизни [Электронный ресурс]. - URL: <https://lifeingos.ru/news/cifrovaa-zizn-kak-novye-tehnologii-menaut-strahovanie-zizni>. (дата обращения: 18.04.2022).
48. Дистанционные сервисы: как цифровые технологии меняют страховой рынок [Электронный ресурс]. – URL: <https://volga.news/article/496179.html> (дата обращения: 18.04.2022).
49. Цифровизации страхового бизнеса в России [Электронный ресурс]. – URL: <http://www.sluchay.ru/> (дата обращения: 22.04.2022).
50. Цыганков А.А. Цифровизации страхового рынка: задачи, проблемы и перспективы [Текст] / А.А. Цыганков, Д.В. Брызгалов // Экономика и управление. - М., 2018. – № 2. - с. 111 – 120.
51. Ильин И.В. Цифровая трансформация как фактор формирования архитектуры и ИТ- архитектуры предприятия [Текст] / И.В. Ильин, А.И. Левина,

А.С. Дубгорн // Научный журнал НИУ ИТМО. – Санкт – Петербург, 2019.- №3. – с. 50 – 55.

52. Кудрявцев Д.В. Архитектура предприятия: переход от проектирования ит - инфраструктуры к трансформации бизнеса [Текст] / Д.В. Кудрявцев, М.Ю. Арзуманян // Журнал менеджмента Russian Management Journal. – Санкт – Петербург, 2017. - № 2. – с. 193 -224.

53. Мобильное решение для страховых компаний [Электронный ресурс]. - URL: <https://www.simbirsoft.com/company/news/simbirsoft-vypustila-mobilnoe-reshenie-dlya-strakhovykh-kompaniy> (дата обращения: 18.04.2022).

54. Гоменюк Н. В Концептуальные принципы оптимизации работы предприятия на основе архитектурного подхода [электронный ресурс] / В.Н. Гоменюк // Вестник института экономических исследований - 2017. - URL: <https://cyberleninka.ru/article/n/kontseptualnye-printsipy-optimizatsii-raboty-predpriyatiya-na-osnove-arhitekturnogo-podhoda> (дата обращения: 20.05.2022)

55. Громов Г. Г. Национальные информационные ресурсы: проблемы промышленной эксплуатации [Текст] / Г. Г. Громов. – М.: Наука, – 240 с. (дата обращения: 15.05.2022).

56. Уэбстер Ф. Д. Теории информационного общества [Текст] / Ф. Уэбстер. – М.: Аспект-Пресс, 2017. – 398 с.

57. Онлайн – платформа Microsoft Azure [электронный ресурс]. – URL: <https://studio.azureml.net/Home/ViewWorkspaceCached/55bd3e40055748df9f53bcc9adbabeab?#Workspace/TrainedModels/ListTrainedModels> (дата обращения: 27.04.2022).

58. Цифровой полис. Как технологии меняют страхование [Электронный ресурс]. – URL: <http://republic.ru/posts/91243> (дата обращения: 05.05.2022).

59. Цифровая трансформация бизнеса: как и зачем меняться в digital-эру [Электронный ресурс]. – URL: <https://vc.ru/flood/42092-cifrovaya-transformaciya-biznesa-kak-i-zachem-menyatsya-v-digital-eru> (дата обращения: 10.05.2022).

60. Официальный сайт ПАО СК «Росгосстрах» [электронный ресурс]. – URL: <https://www.rgs.ru> (дата обращения: 27.04.2022).

61. Коплунова Н.В. Концепция развития информационных ресурсов. [Текст] / Н.В. Каплунова, В.В.Шарыхин, С.В.Хмельницкий / Под ред. С.В.Хмельницкого. – СПб.: Европейский университет в Санкт-Петербурге, 2017. – 168 с.

62. Федюлин А.А. Информационные технологии [Текст] / А.А. Федюлин. - М.: КноРус, 2014. - 472 с.

63. Костромин В.И. Управление информационными ресурсами предприятия [Электронный ресурс] / В.И. Костромин // Экономика и менеджмент предприятия – 2010. – URL: <https://cyberleninka.ru/article/n/upravlenie-informatsionnymi-resursami-predpriya> (дата обращения: 03.05.2022).

64. Гумерова Г.Р. Информационные ресурсы: теория и практика управления [электронный ресурс] / Г.Р. Гумерова // Актуальные проблемы экономики и права – 2011. - URL: <https://cyberleninka.ru/article/n/informatsionnye-resursy-teoriya-i-praktika> (дата обращения: 26.05.2022).

65. Авдонин Л.А. Управление информационными ресурсами корпоративной сети организации [электронный ресурс] / Л.А. Авдонин, В.В. Смогунов, Д.П. Грузин // Информационные технологии и системы – 2007. – URL: <https://cyberleninka.ru/article/n/upravlenie-informatsionnymi-resursami-korporativnoy-seti-organizatsii> (дата обращения: 25.04.2022).

66. Крупнейшие страховщики продолжают наращивать агентскую сеть [Электронный ресурс]. - URL: <http://www.asn-news.ru/news/71354> (дата обращения: 05.05.2022).

67. Кудрявцев Д.В. Разработка учебной методологии управления архитектурой предприятия [Электронный ресурс] / Д.В. Кудрявцев, Е.П. Заременских, М.Ю. Арзуманян // Открытое образование - 2017. - URL: <https://cyberleninka.ru/article/n/razrabotka-uchebnoy-metodologii-upravleniya-arhitekturoy-predpriyatiya> (дата обращения: 28.04.2022).

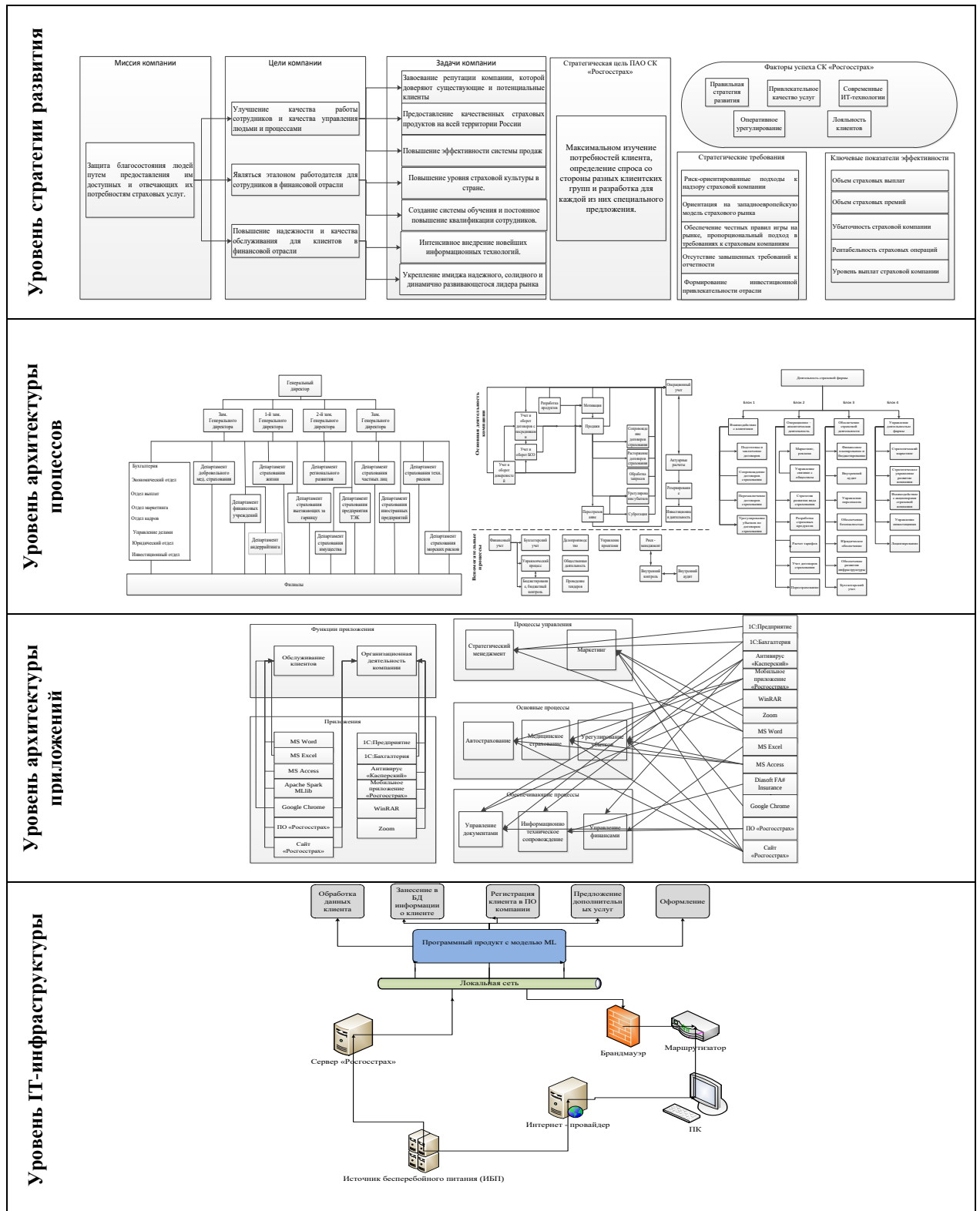
68. Григорьева А.А. Место и роль бизнес-модели в управлении архитектурой предприятия [Электронный ресурс] / А.А., Григорьева // Экономика и бизнес – 2018. - URL: <https://cyberleninka.ru/article/n/mesto-i-rol-biznes-modeli-v-upravlenii-arhitekturoy-predpriyatiya-na-primere-sbytovyh-strukturnk> (дата обращения: 22.05.2022).

69. Тесля А.А. Применение архитектурного подхода к предприятию здорового питания [Электронный ресурс] / А.А. Тесля // Вопросы студенческой науки – 2017. – URL: <https://cyberleninka.ru/article/n/primenenie-arhitekturnogo-podhoda-k-predpriyatiyu-zdorovogo-pitaniya> (дата обращения: 02.05.2022).

70. Маклаков С.В. Моделирование бизнес-процессов с BPMN [Текст] / С.В. Маклаков. – М.: ДИАЛОГМИФИ, 2018.

ПРИЛОЖЕНИЕ А

Полная модель архитектуры ПАО СК «Росгосстрах»⁴⁹



⁴⁹ Составлено автором в MS Visio