

БИЗНЕС-ПЛАН ИННОВАЦИОННОГО ПРОЕКТА

(Пример, система анализа мультимодальных распределенных данных)

1. НАЗВАНИЕ ПРОЕКТА

Платформа анализа распределенных данных / Distributed Data Analysis Platform

Вариант внедрения: система анализа мультимодальных распределенных данных.

Область применения: финансовые рынки и страхование, федеральные и муниципальные государственные структуры, образование, умные квартиры/дома/кварталы/города, медицина, кибербезопасность и др.

2. НАУЧНАЯ СОСТАВЛЯЮЩАЯ ИННОВАЦИОННОГО ПРОЕКТА

2.1. Научная новизна и обоснование предлагаемых в проекте решений.

В ходе выполнения проекта будет проводиться разработка и программная реализация методов анализа мультимодальных данных, размещенных на разных распределенных источниках. В основе таких методов лежит технология федеративного обучения. Полученные результаты будут реализованы в виде прототипа системы анализа мультимодальных данных на распределенных источниках и апробированы на реальных данных. Система будет использовать ранее разработанную коллективом библиотеку ФО для Java (FL4J)

Технология ФО была предложена компанией Google Inc в 2017. На ее основе в настоящее время ведутся разработки библиотек как с открытым исходным кодом, так и проприетарные библиотеки.

Исследование перечисленных библиотек ФО [1], показало, что ни одно из существующих решений в области ФО не позволяет работать с мультимодальными данными на разных источниках. Это существенно снижает возможности машинного обучения и подтверждает актуальность предлагаемого проекта.

Новизной предлагаемого решения является развитие технологии ФО для работы с мультимодальными данными на разных источниках и программная реализация в библиотеке FL4J. В частности:

- методов прямого и обратного преобразования моделей машинного обучения в унифицированный формат глобальной модели, будет отличаться от существующих возможностью привести к единому формату разные модели и комбинировать их в единую, с учетом характеристик данных, на которых они обучены;
- методов комбинирования разных моделей машинного обучения, полученных на различных источниках, отличающихся от существующих мультимодальностью комбинируемых моделей;
- методов ФО для мультимодальных данных, полученных за единый интервал времени об одних и тех же объектах и явлениях, отличающихся от существующих мультимодальностью данных на разных источниках, на которых будет осуществляться обучение;
- методов применения глобальной модели к мультимодальным данным на разных источниках, в отличие от существующих, позволяющая получать единый результат (оценку ситуации, прогноз и т.п.) с учетом всех доступных мультимодальных данных.

Реализация данных методов существенно повысит возможности применения машинного обучения, т.к. снимет ограничения на анализ данных об одних и тех же объектах и явлениях как имеющих разный формат, так и находящихся на разных источниках и не передаваемых в силу разных причин 3-ей стороне (например, в облако).

2.2. Создаваемый коммерческий продукт и его характеристики.

В результате выполнения проекта предполагается разработка и программная реализация методов анализа мультимодальных (разнородных: видео, аудио, текст, структурированные и др.) данных, размещенных на разных распределенных источниках. В основе таких методов лежит технология федеративного обучения. Такие методы позволят существенно повысить качество машинного обучения за счет использования всей доступной информации об объектах или явлениях.

Конечным продуктом проекта будет прототип системы анализа мультимодальных данных, располагающихся на распределенных источниках, без передачи данных 3й стороне. Система будет использовать разработанную ранее библиотеку федеративного обучения FL4J.

В результате система будет иметь следующие потребительские характеристики:

- работа с конфиденциальной информацией: персональные данные, коммерческая тайна, служебная тайна и т.п.;
- использование каналов связи с ограниченной пропускной способностью в системах, близких к реальному времени (Wi-Fi, 3G, спутниковая связь и т.п.);
- обучение на мультимодальных данных на разных источниках: измерения от датчиков, видео- и аудио-потоки, текстовая информация, структурированная информация из хранилищ данных и т.п.;
- анализ (применение обученных моделей) на мультимодальных данных на разных источниках в реальном времени: измерения сенсоров, видеоаналитика, online данные и т.п.

Примерами прикладных областей, где будут востребованы интеллектуальные системы поддержки принятия решений с такими характеристиками, являются:

- федеральные и муниципальные государственные структуры - анализ безопасности, оценка и прогнозы социального развития, занятости населения и т.п.;
- образование – анализ обучающегося, адаптивное обучение и др.
- умные квартиры/дома/кварталы/ города - повышение безопасности, комфорта населения;
- промышленность - контроль качества, технологических процессов, аварийности и надежности оборудования т.п.
- финансовые рынки – оценка кредитоспособности клиентов, анализ и предсказание изменений котировок ценных бумаг и др.;
- медицина – рекомендации здорового образа жизни, диагностика пациентов, прогнозирование течения болезни и др.;
- страхование – оценка страхового риска, страхование автомобилей и т.п.;
- торговля – оценка потребительского спроса, персонификация предложений и услуг, прогнозирование изменения спроса и др.
- кибербезопасность - выявление инсайдеров, сетевых атак, спам и т.п.

Состав системы:

Система анализа мультимодальных распределенных данных включает серверную часть и клиентские модули.

Данные анализируются в месте их хранения у владельца данных. Клиент федеративного обучения, установленный у владельца данных, проводит анализ данных, выдает результаты анализа в информационную систему владельца данных, получает корректировки и может «доучивать» модель. На сервер при этом передаются только результаты обучения моделей.

Сервис ФО агрегирует мультимодальные модели, после чего новые настройки рассылаются сервером на клиентские модули.

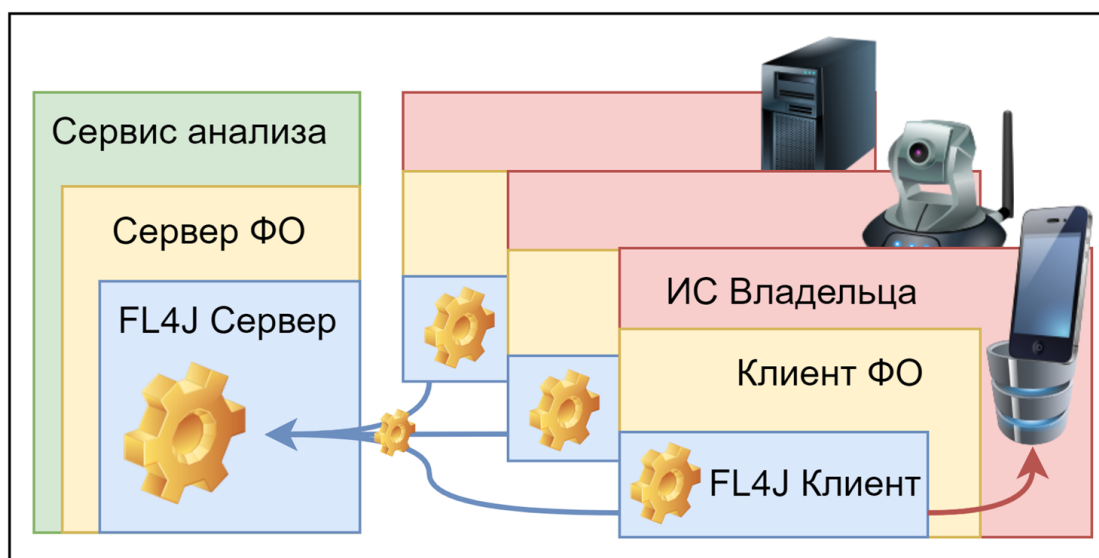


Рис.1. Описание технологии федеративного обучения и ее применения к обработке мультимодальных данных

2.3. Методы и способы решения поставленных задач для получения ожидаемых характеристик.

В научной литературе представлен ряд подходов, позволяющих обучать глобальную модель анализа на мультимодальных данных, принадлежащих разным источникам [2-5]. Однако у них есть ряд недостатков. В частности, они разработаны для обучения определенных моделей анализа, и используют достаточно ресурсоемкие криптографические преобразования для обеспечения конфиденциальности используемых для обучения данных. Использование таких криптопротоколов значительно увеличивает как время обучения, так и время логического вывода. Кроме того, поскольку такие криптопротоколы основаны на большом числе раундов одноранговой (peer-to-peer) связи между участниками обучения, практическое использование таких подходов в системах с плохой связью затруднено.

Перспективным способом обучения модели анализа на мультимодальных данных видится обучение независимых локальных моделей различного типа и их объединение путем обучения глобальной модели на основе параметров локальных моделей. Так, в [6] был предложен подход, который позволяет учитывать различные вычислительные возможности участников ФО, обучать локальные нейронные сети различной архитектуры и объединять их в одну глобальную модель. В [7] представлено развитие этого подхода, в нем для оптимизации взаимодействия между участниками ФО предлагается обмениваться предсказаниями моделей на общем неразмеченном наборе данных, а не параметрами локальных моделей. Предложенные подходы предназначены для решения проблемы обучения на данных, имеющих неоднородное распределение, т.е. Non-IID данные. В Российской федерации проблемами ФО занимается НИУ ВШЭ, наиболее близкой к решаемой задаче является исследование, связанное с построением распределенного обучения глубоких нейронных сетей в условиях неравномерного распределения вычислительных ресурсов у клиентов ФО [8].

В настоящем проекте будет решаться проблема объединения мультимодальных моделей, обученных на мультимодальных данных. Для этого будут решаться следующие научные и научно-технические задачи:

1. разработка способа описания мультимодальных данных на разных источниках в виде

единого набора данных, синхронизированных по интервалу времени и набору объектов и/или явлений, которые они описывают;

2. разработка методов прямого и обратного преобразования моделей машинного обучения в унифицированный формат глобальной модели, описывающий результаты обучения в виде иерархии элементов, представляющих извлеченные из данных закономерности;

3. развитие методов комбинирования разных моделей машинного обучения, полученных на различных источниках;

4. развитие методов ФО для мультимодальных данных, полученных за единый интервал времени об одних и тех же объектах и явлениях (методы ФО для вертикально разделенных данных);

5. развитие методов применения глобальной модели к мультимодальным данным на разных источниках, позволяющих получать единый результат (оценку ситуации, прогноз и т.п.) с учетом всех доступных данных;

6. развитие методов оценки качества глобальной модели на мультимодальных данных на разных источниках;

7. программная реализация методов в библиотеке ФО FL4J [9] для дальнейшего внедрения в интеллектуальные системы поддержки принятия решений;

8. апробация реализованных методов и алгоритмов на реальных мультимодальных данных (собранных и аннотированных вне рамок данного проекта).

Способ описания мультимодальных данных будет представлять собой набор отношений физических атрибутов реальных данных, хранящихся на источниках, и логических атрибутов набора данных, к которому будут применяться методы машинного обучения. В отличие от существующих способов описания, они будут содержать описание атрибутов для неструктурированных данных и способы их извлечения. Таким образом, отношения будут содержать следующую метаинформацию:

- источник местонахождения физических данных;

- способ преобразования физических атрибутов в логические: прямое, объединение и пивот преобразование;

- способ извлечения атрибутов из неструктурированных данных (изображений, текстов, аудио и др.);

- способ синхронизации - приведение к единому для всех данных временному отрезку;

- способ идентификации - определение тождественных объектов и/или явлений, которые данные описывают.

Способ описания мультимодальных данных будет реализован в библиотеке FL4J на базе существующей модели описания данных, что позволит применять к мультимодальным данным из разных источников все алгоритмы библиотеки.

Методы преобразования моделей машинного обучения в унифицированный формат глобальной модели [10], реализованный в библиотеке FL4J. Подобные преобразования позволят представлять разные модели машинного обучения в едином виде и в дальнейшем объединить их. Преобразования будут использовать описание мультимодальных данных на разных источниках в виде единого набора данных.

Методы комбинирования моделей машинного обучения, полученных на мультимодальных данных на разных источниках в глобальную модель. Комбинироваться будут модели, преобразованные в унифицированный формат. При этом агрегация параметров моделей будет осуществляться на основании описания мультимодальных данных на разных источниках.

Методы ФО для мультимодальных данных будут основаны на методах ФО для вертикально распределенных данных и осуществлять диспетчеризацию обучения разных моделей на разных источниках на следующих этапах обучения:

1. описания мультимодальных данных на разных источниках;

2. обучение локальных моделей на каждом из источников;

3. преобразование локальных моделей в унифицированный формат глобальной модели;

4. сбор всех локальных моделей на сервере ФО;

5. комбинирование преобразованных моделей в глобальную модель;
 6. рассылка комбинированной глобальной модели на клиенты ФО;
 7. обратное преобразование комбинированной глобальной модели в локальные модели машинного обучения;
 8. оценка качества глобальной модели ;
 9. повторение шагов 2-8 до завершения обучения или пока поступают новые данные.
- В отличие от существующих методов ФО новые методы будут использоваться в том числе и для неструктурированных данных разного формата: изображения, аудио, текста и др. Методы применения глобальной модели к мультимодальным данным на разных источниках будут осуществлять локальный анализ новых мультимодальных данных на разных источниках и объединять полученные от них результаты. Методы будут выполнять следующие общие этапы:

1. отправка глобальной модели на клиенты ФО;
2. обратное преобразование глобальной модели в локальные модели машинного обучения;
3. применение моделей машинного обучения к новым данным;
4. отправка результатов применения моделей на локальных новых данных на сервер ФО;
5. формирование единого результата применения глобальной модели.

Программная реализация разработанных методов будет выполнена в библиотеке ФО FL4J, которая интегрирована с библиотекой глубокого обучения DL4J и содержит ряд “классических” алгоритмов машинного обучения. Библиотека FL4J реализует ряд уникальных решений, которые будут основой в разрабатываемых методах:

- блоковое представление алгоритмов машинного обучения, которое позволяет переносить блоки алгоритма, выполняющие вычисления над данными на источники и выполнять их параллельно как по данным, так и по задачам;
- описание информации о данных в виде набора отношений физических атрибутов реальных данных, хранящихся на источниках, и логических атрибутов набора данных, к которому будут применяться методы машинного обучения, а также способов преобразования физических атрибутов в логические: прямое, агрегацию и пивот преобразование;
- метод трансформации алгоритмов машинного обучения как для горизонтального, так и для вертикального распределения данных, основанный на операциях из области оптимизаций компиляции, таких как перестановка и расщепление циклов (loop interchange и loop fission) [11], что позволяет практически любой алгоритм машинного обучения применить к вертикально разделяемым данным.
- унифицированную модель, описывающую результаты обучения в виде иерархии элементов, представляющих извлеченные из данных закономерности.

Апробация полученных результатов будет осуществляться на реальных данных о перемещении наземного транспорта и включающих в себя: данные от сенсоров (акселерометра, гироскопа, магнитометра), видеокамеры, микрофона, GPS-приемника, а также внешней информации о внешней среде (погодных условиях и др). Данные будут предоставлены компанией ООО “Смартилайзер Рус” в аннотированном виде. Суммарный объем записанных и аннотированных данных около 150 часов или около 500 миллионов отсчетов сенсоров. Кроме того, компания ООО “Смартилайзер Рус” предоставит программные средства для просмотра и аннотирования данных, для формирования очередей (под-наборов) для обучения и тестирования моделей и для просмотра результатов экспериментов, что позволит пополнить тестовый набор в короткие сроки.

2. R. Xu, N. Baracaldo, Y. Zhou, A. Anwar, J. Joshi, and H. Ludwig. 2021. FedV: Privacy-Preserving Federated Learning over Vertically Partitioned Data. In Proceedings of the 14th ACM Workshop on Artificial Intelligence and Security (AISeC '21). Association for Computing Machinery, New York, NY, USA, 181–192. <https://doi.org/10.1145/3474369.3486872>
3. Q. Zhang, B. Gu, C. Deng, and H Huang. Secure bilevel asynchronous vertical federated learning with backward updating. arXiv preprint arXiv:2103.00958, 2021.

4. C. Wang, J. Liang, M. Huang, B. Bai, K. Bai, and H. Li. Hybrid differentially private federated learning on vertically partitioned data. arXiv preprint arXiv:2009.02763, 2020.
5. T. Chen, X. Jin, Y. Sun, and W. Yin. VafL: a method of vertical asynchronous federated learning. arXiv preprint arXiv:2007.06081, 2020
6. D. Enmao, J. Ding and V. Tarokh. HeteroFL: Computation and Communication Efficient Federated Learning for Heterogeneous Clients.” ArXiv abs/2010.01264 (2021): n. pag.
7. Y.J.Cho, J. Wang, T. Chiruvolu, & G. Joshi, G. (2021). Personalized Federated Learning for Heterogeneous Clients with Clustered Knowledge Transfer. ArXiv, abs/2109.08119.
8. Diskin, Michael et al. “Distributed Deep Learning in Open Collaborations.” NeurIPS (2021). ArXiv, abs/2106.10207.
9. Kholod I. I. et al. FL4J—Federated Learning Framework for Java //Intelligent Distributed Computing XIV. – С. 225. DOI: 10.1007/978-3-030-96627-0_21
10. Ivan I. Kholod, Andrey V. Shorov. Unification of Mining Model for Parallel Processing. In: Proceeding of 2017 IEEE North West Russia Section Young Researchers in Electrical and Electronic Engineering Conference. (2017 ElConRusW), pp. 450–455. IEEE Xplore (2017). 10.1109@ElConRus.2017.7910588
11. Ivan Kholod, Andrey Shorov, and Sergei Gorlatch. Improving Data Mining for Differently Distributed Data in IoT Systems // The 13th International Symposium on Intelligent Distributed Computing (IDC 2019) pp 75-85 DOI: 10.1007/978-3-030-32258-8_9

2.4. Имеющийся у коллектива предприятия научный задел по предлагаемому НИОКР

Библиотека ФО (Federated Learning for Java - FL4J) <https://gitlab.fkti.etu.ru/fl4j/fl4j-framework> - реализующая алгоритмы федеративного обучения, как в симуляционном, так и в реальном режиме, интегрированная с библиотекой Deep Learning for Java, для применения методов глубокого обучения. Библиотека FL4J реализует ряд уникальных решений, которые будут основой в разрабатываемых методах:

- блоковое представление алгоритмов машинного обучения, которое позволяет переносить блоки алгоритма, выполняющие вычисления над данными на источники и выполнять их параллельно как по данным, так и по задачам;
- описание информации о данных в виде набора отношений физических атрибутов реальных данных хранящихся на источниках и логических атрибутов набора данных к которому будут применяться методы машинного обучения, а также способов преобразования физических атрибутов в логические: прямое, агрегацию и пивот преобразование;
- метод трансформации алгоритмов машинного обучения как для горизонтального, так и для вертикального распределения данных, основанный на операциях из области оптимизаций компиляции, таких как перестановка и расщепление циклов (loop interchange и loop fission) [2], что позволяет практически любой алгоритм машинного обучения применить к вертикально разделяемым данным.
- унифицированную модель, описывающую результаты обучения в виде иерархии элементов, представляющих извлеченные из данных закономерности.

Библиотека FL4J соответствует 4-му уровню технологической готовности (УТГ) по ГОСТ Р 56861-2016: реализованы базовые части основного программного фреймворка, приняты решения в области набора программных технологий для реализации сервиса. Ядро и алгоритмы библиотеки FL4J распространяются под лицензией Apache 2.0. Это позволяет привлекать сообщество разработчиков к расширению состава алгоритмов. Лицензией Apache 2.0 разрешает: Коммерческое использование; Распространение; Изменение; Личное использование; Предоставление патентных прав Лицензией Apache 2.0 требует: Упоминания авторства и лицензии в работе; Указывать изменения, внесенные в работу. От пользователей она требует только, если работа была изменена, писать об этом, и указывать исходное авторство. Лицензия отдельно оговаривает,

что для производных работ нельзя использовать те же названия, если они являются торговыми марками.

Компания ООО Смартлайлез Рус для выполнения проекта предоставляет:

- Стенд для изучения и сравнения фреймворков федеративного обучения, соответствующий 8-му уровню УТГ (прошел апробацию у заказчика);
- Набор аннотированных данных: около 150 часов или около 500 миллионов отсчетов сенсоров;
- Программные средства для просмотра и аннотирования данных, для формирования очередей (под-наборов) для обучения и тестирования моделей и для просмотра результатов экспериментов, соответствующие 7-му уровню УТГ (прошли апробацию в ООО Смартгайзер Рус).

Основные результаты описаны в следующих ключевых публикациях:

1. Novikova, E.; Fomichov, D.; Kholod, I.; Filippov, E. Analysis of Privacy-Enhancing Technologies in Open-Source Federated Learning Frameworks for Driver Activity Recognition. *Sensors* 2022, 22, 2983. <https://doi.org/10.3390/s22082983>
2. Efremov, M.A., Kholod, I.I., Kolpaschikov, M.A.: Java federated learning framework architecture. In: 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), pp. 306–309 (2021). DOI 10.1109/ElConRus51938.2021.9396508
3. Kholod, I.; Yanaki, E.; Fomichev, D.; Shalugin, E.; Novikova, E.; Filippov, E.; Nordlund, M. Open-Source Federated Learning Frameworks for IoT: A Comparative Review and Analysis. *Sensors* 2021, 21, 167. <https://doi.org/10.3390/s21010167>
4. Ivan Kholod, Andrey Shorov, and Sergei Gorlatch. Efficient Distribution and Processing of Data for Parallelizing Data Mining in Mobile Clouds. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, 11(1):2-17, Mar. 2020
DOI:10.22667/JOWUA.2020.03.31.002
5. Kholod, I., Rukavitsyn, A., Paznikov, A. et al. Parallelization of the self-organized maps algorithm for federated learning on distributed sources. *J Supercomput* (2020).
<https://doi.org/10.1007/s11227-020-03509-2>
6. M. A. Efremov and I. I. Kholod, "Architecture of Swarm Robotics System Software Infrastructure," 2020 9th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 2020, pp. 1-4, doi: 10.1109/MECO49872.2020.9134247.
7. Ivan Kholod, Andrey Shorov, and Sergei Gorlatch. Improving Data Mining for Differently Distributed Data in IoT Systems // The 13th International Symposium on Intelligent Distributed Computing (IDC 2019) pp 75-85
DOI: 10.1007/978-3-030-32258-8_9
8. Kholod I. I. et al. FL4J—Federated Learning Framework for Java //Intelligent Distributed Computing XIV. – С. 225. DOI: 10.1007/978-3-030-96627-0_21

2.5. Планы по созданию и защите интеллектуальной собственности, в том числе по подаче международных заявок в соответствии с договором о патентной кооперации (РСТ).

На элементы библиотеки и реализованные в ней решения и алгоритмы оформлены документы, подтверждающие права на интеллектуальную собственность:

1. Холод И.И., Малов А.В., Родионов С.В. Способ распараллеливания интеллектуального анализа данных в вычислительной среде. // Патент на изобретение №2745018 от 18 марта 2021 г
2. Ефремов М.А., Колпашиков М.А., Табаков П.Л. Программный адаптер для управления сервером федеративного обучения. //Свид. о государств. Регистрации программы для ЭВМ № 2021669639 от 23.11.2021.
3. Ефремов М.А., Табаков П.Л. Программа для регистрации клиентов федеративного

- обучения. //Свид. о государств. Регистрации программы для ЭВМ № 2021669370 от 23.11.2021.
4. Ефремов М.А., Аристархов И.Е. Программа управления клиентом федеративного обучения. //Свид. о государств. Регистрации программы для ЭВМ № 2021669391 от 23.11.2021.
 5. Холод И.И. Программа для распараллеливания по данным процессов интеллектуального анализа. //Свид. о государств. Регистрации программы для ЭВМ № 2016610768 от 19.01.2016.
 6. Холод И.И. Программа подготовки набора функциональных блоков интеллектуального анализа данных к параллельному выполнению. //Свид. о государств. Регистрации программы для ЭВМ № 2016610772 от 19.01.2016.
 7. Холод И.И. Функционально-блочная программа построения одноатрибутных классификационных правил. //Свид. о государств. Регистрации программы для ЭВМ № 2015611447 от 29.01.2015.
 8. Холод И.И. Функционально-блочная программа параллельного построения одноатрибутных классификационных правил. //Свид. о государств. Регистрации программы для ЭВМ № 2015611451 от 29.01.2015.
 9. Холод И.И. Функционально-блочная программа параллельного поиска ассоциативных правил. //Свид. о государств. Регистрации программы для ЭВМ № 2014660763 от 23.10.2014.
 10. Холод И.И. Блоковая структура выполнения алгоритма поиска ассоциативных правил AprioriTID. //Свид. о государств. Регистрации программы для ЭВМ № 2014618030 от 12.08.2014.
 11. Холод И.И. Программа блоковой декомпозиции алгоритма кластеризации kmeans. //Свид. о государств. Регистрации программы для ЭВМ № 2012610929 от 20.01.2012.
 12. Холод И.И. Программа декомпозиции алгоритмов интеллектуального анализа данных. //Свид. о государств. Регистрации программы для ЭВМ № 2012610930 от 20.01.2012.
 13. Холод И.И., Каршиев З.А. Блоковая структура выполнения алгоритма классификации Naïve Bayes. //Свид. о государств. Регистрации программы для ЭВМ № 2012660852 от 29.11.2012.
 14. Холод И.И., Накидкин А.В. Блоковая структура для параллельного выполнения алгоритмов интеллектуального анализа данных. //Свид. о государств. Регистрации программы для ЭВМ № 2012660853 от 29.11.2012.

Поданы заявки на регистрацию имеющихся компонентов фреймворка федеративного обучения как программ для электронных вычислительных машин и баз данных:

Программа для управления клиентом владельца наборов данных ver. 2.0. Авторы: Ефремов Михаил Александрович, Колпащиков Максим Алексеевич, Сидорова Полина Александровна, Петухов Владимир Дмитриевич

Программа для регистрации клиентов федеративного обучения ver. 2.0. Авторы: Ефремов Михаил Александрович, Колпащиков Максим Алексеевич, Сидорова Полина Александровна, Петухов Владимир Дмитриевич

Программа для анализа данных федеративным методом. Авторы: Ефремов Михаил Александрович, Колпащиков Максим Алексеевич, Сидорова Полина Александровна, Петухов Владимир Дмитриевич

Мобильная программа управления клиентом владельца наборов данных. Авторы: Ефремов Михаил Александрович, Забалуев Даниил Антонович

Имеется лицензионный договор с СПбГЭТУ ЛЭТИ на использование, доработку и коммерциализацию компонентов фреймворка федеративного обучения.

ООО «Смартилайзер рус» планирует зарегистрировать разработанное в проекте программное обеспечение в Федеральном государственном бюджетном учреждении «Федеральный институт промышленной собственности» (ФИПС) и получить свидетельства о государственной регистрации программ для электронных вычислительных машин и баз данных.

Настройки нейронных сетей, реализующих модули формирования метрик и прогнозов, планируется сохранять как ноу-хау компании.

3. СООТВЕТСТВИЕ ПРОЕКТА ВЫСОКОТЕХНОЛОГИЧНОМУ НАПРАВЛЕНИЮ И НАПРАВЛЕНИЮ РАЗРАБОТКИ ОТЕЧЕСТВЕННОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Основные направления программы: «Цифровые технологии», «Искусственный интеллект».

Поднаправление: Системы обработки и хранения информации. Инструменты для анализа больших данных (Big Data).

Фокусная тематика: Инструменты для анализа больших данных (Big Data)

Приоритетные направления: Информационно-телекоммуникационные системы

Предполагаемый результат: «Платформа анализа мультимодальных данных из разных источников на основе технологии ФО, которая позволяет взаимно “обогащать” данные без их передачи кому-либо и существенно уменьшить стоимость владения большими данными»

Отрасли экономики и социальной сферы, в которых реализуется проект: финансовые рынки и страхование, федеральные и муниципальные государственные структуры, образование, умные квартиры/ дома/ кварталы/ города, медицина, кибербезопасность и др.

Проект соответствует приоритетным направлениям поддержки разработки отечественного программного обеспечения в рамках высокотехнологичных направлений и сквозных цифровых технологий: Переход к передовым цифровым, интеллектуальным производственным технологиям, роботизированным системам, новым материалам и способам конструирования, создание систем обработки больших объемов данных, машинного обучения и искусственного интеллекта

Результатами проекта будут программная реализация методов анализа для мультимодальных данных, размещенных на разных источниках без их передачи 3-ей стороне. Таким образом, результаты проекта будут востребованы в областях использующих машинное обучение в случае:

- работы с мультимодальными данными на разных источниках: измерения от датчиков, видео- и аудио-поток, текстовая информация, структурированная информация из хранилищ данных и т.п.;
- работы с конфиденциальными данными (персональные данные, коммерческая тайна, служебная тайна и т.п.);
- использование каналов связи с ограниченной пропускной способностью в системах близких к реальному времени (Wi-Fi, 3G, спутниковая связь и т.п.);
- необходимости выполнять прогнозирование на основе мультимодальных данных в реальном времени (измерения сенсоров, видеоаналитика, online данные и т.п.).

Примерами прикладных областей, где возможны подобные ситуации и могут быть востребованы полученные в проекте результаты, являются:

- федеральные и муниципальные гос. структуры – анализ безопасности, социального развития, занятости населения;
- образование – анализ обучающегося, адаптивное обучение и др.
- умные квартиры/дома/кварталы/ города – повышение безопасности, комфорта населения;
- промышленность – контроль качества, технологических процессов, аварийности и надежности оборудования т.п.
- финансовые рынки – оценка кредитоспособности клиентов, анализ и предсказание изменения котировок ценных бумаг и др.;
- медицина – рекомендации здорового образа жизни, диагностика пациентов, прогнозирование течения болезни и др.;
- страхование – оценка страхового риска, страхование автомобилей и т.п.;

- торговля – оценка потребительского спроса, персонализация предложений и услуг, прогнозирование изменения спроса и др.
- кибербезопасность - выявление инсайдеров, сетевых атак, спама и т.п.

4. ПЕРСПЕКТИВЫ КОММЕРЦИАЛИЗАЦИИ

4.1. Объем и емкость рынка продукта, анализ современного состояния и перспектив развития отрасли, в которой реализуется инновационный проект (если рынок новый, необходимо представить его описание).

Практически каждый сегмент IT рынка или уже тесно связан с обработкой данных, или к этому неизбежно стремится.

По отчету компании IBM количество хранимых данных удваивается каждые два года. Основными отраслями, показывающими рост использования аналитики больших данных, согласно отчету SaaS Scout, являются: медицина и фармацевтика, банковское дело, медиарынок (сервисы потокового воспроизведения, социальные сети, игровые сервисы), рынок ритейла, электроэнергетика, строительная отрасль. Анализ больших данных позволяет уменьшить издержки и сократить стоимость для конечного потребителя, получив рыночные преимущества, например, в здравоохранении на 17%, согласно исследованию McKinsey. По данным экспертов IDC, в 2021 году объем мирового рынка больших данных и бизнес-аналитики (BDA) составил \$215,7 млрд, увеличившись на 10,1% относительно 2020-го. В своих расчетах аналитики учитывают коммерческие закупки оборудования, программного обеспечения и услуг, связанных с BDA. Примерно треть расходов на большие данные и бизнес-аналитику в 2021 году пришлось на три отрасли: банковский сектор, дискретное производство и профессиональные услуги. Следующие три по размеру инвестиций в BDA сегмента - непрерывное производство, телеком и правительство - добавили рынку выручку в \$47 млрд по итогам 2021 года.

Ожидается, что мировой рынок машинного обучения (ML) вырастет с 21,17 млрд долларов в 2022 году до 209,91 млрд долларов к 2029 году при среднегодовом темпе роста в 38,8% в прогнозируемый период. (<https://www.fortunebusinessinsights.com/machine-learning-market-102226>)

В статье 2020 года в журнале Nature “The future of digital health with federated learning” рассматривается федеративная технология машинного обучения как способ избежать проблем с безопасностью чувствительных медицинских данных, а сами исследования в области федеративного обучения будут активно проводиться ближайшие 10 лет. Согласно исследовательскому отчету “Global Federated Learning Solutions Market by Application (Drug Discovery, Industrial IoT), Vertical (Healthcare & Life Sciences, BFSI, Manufacturing, Retail & e-Commerce, Energy & Utilities), and Region - Forecast to 2028” («Рынок федеративного обучения по приложениям (...) и регионам — глобальный прогноз до 2028 года»), опубликованному MarketsandMarkets™ в начале 2022 года, в соответствии со сценарием AS-IS, размер глобального рынка федеративного обучения вырастет со 127 миллионов долларов США в 2023 году до 210 миллионов долларов США к 2028 году при совокупном годовом темпе роста (CAGR) 10,6% в течение прогнозируемого периода. (<https://www.marketsandmarkets.com/PressReleases/federated-learning-solutions.asp>)

Отечественный рынок пока выглядит скромнее, но имеет потенциал к взрывному росту. Согласно данным, приведенным Ассоциацией участников рынка больших данных, объем рынка Big Data в России в 2020 году составил 10-30 млрд руб. При этом, в соответствии с усредненными прогнозами отечественных и иностранных экспертов, предполагается рост этого показателя в 10 раз – до отметки 300 млрд руб. к 2024 году.

По данным IDC Worldwide Artificial Intelligence Spending Guide, российский рынок искусственного интеллекта в 2020 году достиг 291 млн долларов США. В 2020 году наблюдался значительный рост инвестиций со стороны государственных организаций, который до 2024 года продолжится со средним ежегодным темпом 26,4%. Этот рост будет поддерживать заявленная государственная программа в области развития искусственного интеллекта (<https://www.idc.com/getdoc.jsp?containerId=prEUR247642121>). Таким образом в 2023 году объем рынка может достичь 588 млн долларов США.

4.2. Конкурентные преимущества создаваемого продукта, сравнение технико-экономических характеристик с отечественными и мировыми аналогами.

В настоящее время широко используются облачные сервисы анализа данных от IT лидеров, такие как Amazon Elastic Mapreduce, Google BigQuery, Apache Spark. Все они используют программные средства, реализующие концепцию MapReduce. Они предполагают, что данные будут передаваться в облако, где выполняется анализ. Это существенно снижает возможность их использования по следующим причинам:

- необходимо передавать данные 3-ей стороне;
- использовать защищенные каналы связи с хорошей пропускной способностью;
- постоянно актуализировать данные в облаке, чтобы анализ проводился на “свежих” данных.

Для преодоления данных проблем компанией Google Inc, предложена в 2017 году концепция ФО. На ее основе в настоящее время ведутся разработки библиотек с открытым исходным кодом:

- TensorFlow Federated (TFF) - от компании Google Inc (США)
<https://github.com/tensorflow/federated>
 - Federated AI Technology Enabler (FATE) – компания Webank (Китай)
<https://github.com/FederatedAI/FATE>
 - Paddle Federated Learning (PFL) – компания Baidu (Китай)
<https://github.com/PaddlePaddle/PaddleFL>
 - PySyft – проект открытого сообщества OpenMined <https://github.com/OpenMined/PySyft>
 - Flower – проект компании Adap GmbH (Германия) <https://github.com/adap/flower>
- Также разработаны и проприетарные библиотеки
- Nvidia Clara Train SDK – компания Nvidia (США)
<https://ngc.nvidia.com/catalog/containers/nvidia:clara-train-sdk>
 - IBM FL – компания IBM (США) <https://www.ibm.com/blogs/research/2020/08/ibm-federated-learning-machine-learning-where-the-data-is/>
 - Swarm Learning - компания Hewlett Packard Enterprise (США)
<https://github.com/HewlettPackard/swarm-learning>

Был проведен сравнительный анализ представленных выше библиотек и фреймворков ФО в части поддержки различных моделей анализа, типов данных и их распределения между клиентами, наличия возможности обучения с подкреплением. Исследование [1] показало, что только библиотека FATE имеет реальную реализацию работы с вертикально распределенными данными, т.е. с данными, содержащими информацию об одних и тех же объектах и явлениях, но об их разных характеристиках на разных источниках. Однако имеется ряд ограничений, связанных с практическим применением данной библиотеки на практике. В частности, FATE реализует только две модели машинного обучения для работы с вертикальными данными: нейронные сети (алгоритм HeteroNN) и деревья решений (алгоритм HeteroGBDT). Текущая реализация алгоритма HeteroNN может анализировать данные только на двух источниках. Еще одним ограничением FATE является работа только со структурированной информацией. Таким образом, наиболее развитая в настоящее время библиотека ФО не позволяет выполнять обучение на мультимодальных данных на разных источниках. Это существенно снижает возможности машинного обучения и подтверждает актуальность предлагаемого проекта.

Новизной предлагаемого решения является развитие технологии ФО для работы с мультимодальными данными на разных источниках и программная реализация в библиотеке FL4J. Это существенно повысит возможности применения машинного обучения, т.к. снимет ограничения на анализ данных об одних и тех же объектах и явлениях, как имеющих разный

формат, так и находящихся на разных источниках и не передаваемых в силу разных причин третьей стороне (например, в облако).

1. Kholod, I.; Yanaki, E.; Fomichev, D.; Shalugin, E.; Novikova, E.; Filippov, E.; Nordlund, M. Open-Source Federated Learning Frameworks for IoT: A Comparative Review and Analysis. *Sensors* 2021, 21, 167. <https://doi.org/10.3390/s21010167>

4.3. Планируемая стоимость продукта. Расчет себестоимости.

Полная себестоимость разработки и производства интеллектуальной системы анализа распределенных данных по нашим расчетам составляет 144 000 000 руб. за 6 лет.

Полная себестоимость включает:

1) затраты понесенные на этапе выполнения НИОКР, создание MVP и полноценной версии продукта 1.0, а именно, заработную плату основного персонала, заработную плату управленческого и вспомогательного персонала, отчисления от заработной платы, на оплату облачных сервисов и лицензионного программного обеспечения,

2) предполагаемые затраты на сопровождение и доработку системы на протяжении 4,5 лет после выполнение стадии НИОКР, а именно, заработную плату основного персонала, заработную плату управленческого и вспомогательного персонала, отчисления от заработной платы, затраты на приобретение и обновление компьютерного оборудования, на оплату облачных сервисов и лицензионного программного обеспечения, расходы на рекламу и сбыт.

Во всех работах предполагается задействовать сбалансированную по опыту работы команду: опытных исследователей и технических лидеров для анализа и постановки задач, для выполнения наиболее ответственных модулей, а также инженеров и младших инженеров, для адаптации платформы под новых заказчиков и для новых наборов данных.

Такой подход позволит обеспечить стоимость команды из 4-5 человек порядка 600 тыс. руб. в месяц и выполнять проекты типа PoC за 2 месяца при стоимости 1,2 млн. руб., выполнять проекты типа MVP за 8 месяцев при стоимости порядка 7,2 млн. руб. и обеспечивать сопровождение внедренных решений при стоимости 250 тыс. руб. в месяц.

На момент активных продаж будут разработаны тарифы с различными комплектами сервисов и несколько уровней поддержки.

4.4. Целевые сегменты потребителей создаваемого продукта и оценка платежеспособного спроса.

Основываясь на данных, приведенных в п. 4.1, можно спрогнозировать размер рынка ФО в Российской Федерации в 2023 году как 2,54 млн долларов США. Имея серьезный задел в виде разработанной библиотеки ФО, набора программных средств для исследований ФО и начальных наборов данных, новый стартап имеет все возможности для того, чтобы стать лидером рынка в РФ и занять долю рынка от 30% до 50% (0,76 - 1,27 млн долларов США в 2023 году)

Использование систем ФО наиболее актуально в случае:

- работа с конфиденциальными данными (персональные данные, коммерческая тайна, служебная тайна и т.п.);
- использование каналов связи с ограниченной пропускной способностью в системах близких к реальному времени (Wi-Fi, 3G, спутниковая связь и т.п.);
- возможности самообучения моделей на локальных данных (наличие обратной связи от пользователя, подтверждение прогноза и т.п.);
- необходимости выполнять прогнозирование на основе данных в реальном времени (измерения сенсоров, видеоаналитика, online данные и т.п.).

Варианты применения ФО можно разбить на три типа:

- прогнозирование (финансовые рынки, медицина, маркетинг, ритейл и др.);
- контроль сложных объектов (промышленность, безопасность, умные дома/города, беспилотный транспорт и др.);
- оценка клиентов (медицина, банковская сфера, страхование, ритейл, образование и др.).

Примерами прикладных областей, где возможны подобные ситуации и могут быть востребованы полученные в проекте результаты, являются:

- финансовые рынки – оценка кредитоспособности клиентов, анализ и предсказание изменение котировок ценных бумаг и др;
- страхование – оценка страхового риска, страхование автомобилей и т.п.;
- федеральные и муниципальные гос. структуры - анализ безопасности, социального развития, занятости населения;
- образование – анализ обучающегося, адаптивное обучение и др.
- умные квартиры/дома/кварталы/ города - повышение безопасности, комфорта населения;
- промышленность – контроль качества, технологических процессов, аварийности и надежности оборудования т.п.
- медицина – рекомендации здорового образа жизни, диагностика пациентов, прогнозирование течения болезни и др.;
- торговля – оценка потребительского спроса, персонализация предложений и услуг, прогнозирование изменения спроса и др.
- кибербезопасность – выявление инсайдеров, сетевых атак, спам и т.п.

Ключевыми факторы принятия решения потребителями являются:

- осознание потенциальным заказчиком проблем, связанных с поиском закономерностей по эмпирическим данным, не поддающимся статистической обработке;
- необходимость быстрого принятия решения на основе хранимых данных;
- извлечение знаний из хранимой информации;
- тестирование собственных алгоритмов в определенной вычислительной среде;
- подбор алгоритмов решения прикладных задач машинного обучения;
- подбор вычислительной среды для реализации определенных алгоритмов;
- необходимость получения решения при невозможности раскрытия эмпирических данных и тем более сохранения их в чужом хранилище.

Возьмем для примера страхование. Методы ИИ/МО уже широко используются в этой индустрии, например: <https://1prime.ru/finance/20200902/831980390.html>

Оптимизация тарифов: при оформлении полиса обученная модель с помощью метрик и ретроспективных данных может сказать, стоит ли повысить или понизить тариф для конкретного потребителя; расчет возможной прибыли от клиента.

Оценка рисков в автостраховании: можно учитывать водительский стаж, время года, историю клиента, социальные факторы, ситуацию в городе, а при назначении цены добавлять состояние рынка, уровень зарплат в городе, прогноз оттока клиентов, чтобы стоимость была сбалансированной.

Мошенничество, урегулирование убытков: модели успешно распознают ситуации, когда злоумышленники пытаются обмануть страховую компанию, подстраивая страховой случай.

Можно экономить, обнаруживая мошенников и оставляя их без выплат; для честных же клиентов компании с помощью ИИ могут упростить процедуру выплаты.

Предотвращение страховых случаев: датчики автомобиля (компьютерное зрение) могут обнаруживать рискованное поведение водителя (закрываются глаза, начал писать смс, закурил и так далее) и сигнализировать о поджидающей опасности, существенно снижая количество ДТП со смертельным исходом.

Прогнозирование: анализ и прогнозирование спроса и оттока клиентов, рекомендации других страховых услуг.

Другое: распознавание космических снимков для "умного" агро-страхования; выявление и исключение ошибок, допущенных из-за человеческого фактора.

Контакт с клиентом: голосовые и чат-боты; технология обработки естественного языка позволяет транскрибировать диалоги агентов с клиентами, на которых потом обучается чат-бот, и помогает выделить в документах признаки страхового мошенничества.

(<https://www.tadviser.ru/>

«Статья: InsurTech - Информационные технологии и цифровизация в страховании»)

Киберстрахование: услуга, которая стала набирать популярность только в 2017 году после масштабной эпидемии вирусов-вымогателей.

Аналитики и страховые компании высоко оценивают перспективы внедрения ИИ/МО и активно работают над внедрением:

McKinsey: внедрение ИИ может повысить производительность страховых процессов и сократить операционные расходы до 40% к 2030 году.

Accenture: благодаря технологиям страховщики в 2019-20 годах смогли сократить время на урегулирование убытков до нескольких дней и уменьшить расходы в пределах 10% от всех трат на урегулирование.

<https://www.asn-news.ru/news/72013>

РЕСО-Гарантия: повышение эффективности продаж, в частности, кросс-продаж полисов ОСАГО

Тинькофф Страхование:

- оценки риска клиента (вероятность мошенничества, возможность денежных убытков);
- оценка вероятности пролонгации страхового полиса или вероятности получить суброгацию по ДТП, в которое попал клиент
- автоматическая оценка суммы убытка по фотографиям

СОГАЗ: расчет рисков, урегулирование убытков, анализ медицинских услуг и работа с персоналом.

Зетта Страхование: персонифицированный подход к потребностям клиентов на основе проанализированных данных.

Капитал Лайф Страхование Жизни:

- рекомендации о периодичности диагностики здоровья застрахованных;
- анализ данных клиентов для формирования адресного предложения клиентам;
- принятие решений по урегулированию убытков, обрабатывая данные лечебных учреждений;
- формирование индивидуальных предложений (сервисы для поддержания здоровья, недвижимости, автомобиля, других активов).

Сбербанк страхование: построение моделей страховых тарифов, в том числе индивидуальных, применяются при урегулировании убытков. Во-первых, это выявление страховых мошенников. Во-вторых, сейчас активно развивается направление по принятию автоматического решения по страховым случаям.

ВСК: распознавание речи и документов, с применением элементов искусственного интеллекта

Либерти страхование: внедрение автоматического скоринга, который станет фундаментом для нового этапа интенсификации процессов практически по всем линиям бизнеса компании.

Однако и аналитики, и страховые компании видят серьезные проблемы со сбором достаточного количества качественных данных:

Forbes: для применения технологий ИИ есть два серьезных барьера: высокая зарегулированность страховой отрасли и сбор качественных данных, без которых невозможно эффективное машинное обучение.

Microsoft: Данные многих страховых компаний хранятся в устаревших изолированных системах. И чтобы успешно внедрять инновации, страховым операторам в первую очередь необходимо модернизировать их.

McKinsey: Различные государственные и частные организации объединятся для создания экосистем для обмена данными для различных вариантов использования в рамках общей системы регулирования и кибербезопасности.

Сбербанк страхование: Набирает популярность внедрение удаленных осмотров с использованием мобильных телефонов. После того, как компания накопит достаточно большой объем данных, можно будет учить модели определению степени повреждений и т.д.

Тинькофф Страхование: На успешность использования той или иной модели влияет не столько ее сложность, сколько объем данных. Компания постоянно работает над его ростом, учитывая все новые факторы оценки, как внутренние, так и внешние. Федеративное обучение – это технология, которая позволяет решить или уменьшить проблему Больших Данных.

4.5. Описание бизнес-модели проекта. Производственный план и план продаж.

На первом этапе результаты могут быть коммерциализированы как проекты апробации технологических решений на наборах данных заказчика. Бизнес-модель реализации таких проектов - оплата с фиксированной ценой или повременная оплата (Т&М). Предполагаемая длительность таких проектов - 1-2 месяца в случае проверки концепции, 3-9 месяцев в случае проверки концепции и интеграции технологического решения в окружение заказчика. Состав команды - 4-5 специалистов. Почасовая оплата от 600 - 3000 руб./час. Будет предлагаться продажа лицензий на использование библиотеки алгоритмов FL4J и фреймворка федеративного обучения. Модель распространения будет простой: фиксированная стоимость без ограничения времени и количества пользователей.

Предполагаемый объем продаж за 6 лет составляет 46 выполненных PoC проектов стоимостью 1,2 млн. руб., 15 проектов типа MVP стоимостью 7,2 млн. руб., и сопровождение порядка 10 платформенных решений при стоимости обслуживания 250 тыс. руб. в месяц.

По мере развития платформы, бизнес-модель будет меняться. После реализации функционала, обеспечивающего возможность работы с новыми видами данных, по мере накопления преднастроенных моделей, ресурсы будут предоставляться в аренду. Оплата будет повременная без детальной тарификации, примерная стоимость аренды 1-й модели на 1 месяц 1000 руб.

Клиентам будут предоставляться скидки, зависящие от объема потребляемых услуг. Владельцы данных, предоставляющие свои данные для обучения моделей, будут получать процент от реализации услуг. Будет предлагаться продажа лицензий на использование фреймворка.

Модель распространения будет учитывать ограничения времени и количество пользователей. Будут предлагаться услуги по созданию рабочих мест аналитика и по обеспечению доступа к наборам данных владельцев данных.

После реализации нового функционала и накопления пула клиентов будет введена повременная оплата с детальной тарификацией до 1 сек. Такая модель будет предоставляться и клиентам, и владельцам данных. Будет предлагаться продажа лицензий на использование платформы.

Модель распространения будет учитывать ограничения времени и количество пользователей. Будут предлагаться услуги по созданию рабочих мест аналитика и по обеспечению доступа к наборам данных владельцев данных. Будут предлагаться услуги по подбору общедоступных данных, привлечению таких данных для обучения моделей с целью улучшения качественных и количественных показателей моделей.

4.6. Стратегия продвижения продукта на рынок.

Реализация стратегии по продвижению продукта стартует по итогам завершения научно-исследовательского этапа с возможностью представления прототипа продукта (MVP).

Во-первых, планируется зарегистрировать стартап в качестве участника Сколково и через мероприятия Сколково выходить на потенциальных клиентов и инвесторов. Во-вторых, планируется работать с профессиональными ассоциациями, такими как ассоциация участников финансового рынка ФИНТЕХ, Российский союз автостраховщиков, ассоциация потребительских кооперативов, ассоциация разработчиков программного обеспечения Руссофт и рядом других ассоциаций и объединений. В-третьих планируется работать с городским правительством, комитетами. Работа во всех указанных направлениях будет направлена как на сбор бизнес-требований к разрабатываемому продукту, на поиск и выполнение пилотных проектов, так и на выявление потенциальных клиентов, заинтересованных как в анализе данных, так и в предоставлении своих данных для анализа.

5. КОМАНДА ПРОЕКТА

5.1. Количество сотрудников, в том числе занятых выполнением НИОКР, направление их деятельности и их квалификация с приложением подтверждающих документов.

Таблица 3. Команда проекта

п/п	ФИО сотрудника	Тип трудоустройства (в штате/по совместительству)	Роль в команде проекта	Квалификация	Опыт реализации проектов по схожей тематике (разработка/коммерциализация)
1	Филиппов Евгений Васильевич	в штат	Менеджер проекта (0,5 ставки)	к.т.н., доцент Руководство проектами, сбор и оформление системных требований, подготовка контрактов, участие в тендерах	35 лет опыта, Диплом к.т.н., Диплом доцента. Резюме с описанием реализованных проектов приложено.
2	Холод Иван Иванович	по совместительству	Научный руководитель (0,5 ставки)	д.т.н., профессор Распределенные вычисления, федеративное машинное обучение	25 лет опыта, Диплом д.т.н., Диплом профессора

3	Банников Алексей Александрови ч	в штат	Системный архитектор (0,5 ставки)	Разработка ПО и архитектуры систем в различных проектах	25 лет опыта, диплом специалиста (Вычислительные машины, комплексы, системы и сети)
4	Новикова Евгения Сергеевна	по совместит ельству	Ведущий инженер по безопасности (0,5 ставки)	к.т.н. Методы защиты персональных данных в системе федеративного обучения	11 лет опыта, Диплом к.т.н.
5	Ефремов Михаил Александрови ч	по совместит ельству	Старший программист бек-энд (0,5 ставки)	Разработка методов и их программная реализация, разработка методик проверки, проведение испытаний	5 лет опыта, Диплом магистра, направление «Программная инженерия»
6	Колпащиков Максим Алексеевич	в штат	Программис т (бэк-энд)	Программная реализация методов, разработка кода макета (back-end)	3 года опыта, Диплом бакалавра, направление “Информатика и вычислительная техника”
7	Сидорова Полина Александровн а	в штат	Программис т ИИ	Подготовка наборов данных, проведение испытаний	2+ года опыт разработки в области МО Диплом бакалавра, направление “Информатика и вычислительная техника”
8	Авдей Светлана Павловна	в штат	Старший программист (front-end)	Разработка методов и их программная реализация, разработка кода макета (front-end)	15 лет опыта, Диплом специалиста, направление “Вычислительные машины, комплексы, системы и сети”
9	Постаногова Светлана Владимировна	в штат	Менеджер по маркетингу и коммуникац иям (0,5 ставки)	Маркетинговые исследования, координация внешних и внутренних коммуникаций проекта.	20 лет опыта, Диплом специалиста, ПГУ.
10	Золотарева Наталья Игоревна	в штат	Ведущий инженер по тестировани ю,	Руководство группой тестирования, разработка тест-кейсов, автоматизация тестирования	10 лет опыта, Диплом (Гурова) специалиста “Автоматизированн ые системы

			автоматизатор		обработки информации и управления”
11	Фомичев Дмитрий Алексеевич	в штат	Программист (ИИ)	Разработка моделей федеративного обучения, проведение испытаний	3+ года опыт разработки в области МО Диплом магистра, направление “Информатика и вычислительная техника”
12	Гурова Екатерина Игоревна	в штат	Инженер по тестированию	Разработка тест-кейсов, ручное тестирование	10+ лет тестирования ПО, Диплом инженера, специальность «Электромеханика»
13	Медведев Евгений Романович	в штат	Программист (ИИ)	Разработка моделей федеративного обучения, проведение испытаний	3+ года опыт разработки в области МО Диплом магистра, направление “Информатика и вычислительная техника”

5.2. Опыт команды в реализации подобных проектов.

Членами научного коллектива в рамках работ, проводимых на факультете Компьютерных Технологий и Информатики ЛЭТИ, выполнены следующие работы по аналогичной тематике:

- разработана концепция построения и адаптации алгоритмов машинного обучения для анализа распределенных данных без их передачи от источников;
- разработана библиотека алгоритмов машинного обучения для анализа распределенных данных в федеративном режиме;
- исследованы существующие системы федеративного обучения;
- разработаны базовые модули фреймворка федеративного обучения, позволяющие быстрое прототипирование и создание MVP под конкретные требования партнеров.

Концепция построения и адаптации алгоритмов машинного обучения для анализа распределенных данных включает в себя:

- представление модели, которая описывает найденные в данных закономерности в виде леса деревьев унифицированных элементов и определяет функции, обеспечивающие их параллельное построение в общей и распределенной памяти;
- функциональную модель алгоритмов машинного обучения, расширяющую теорию λ -исчислений и представляющая его в виде композиции унифицированных потокобезопасных функций, которая может быть перестроена в зависимости от условий выполнения алгоритма;
- метод оптимизации структуры алгоритма машинного обучения, трансформирующий алгоритм в зависимости от типа распределения данных;

- модель параллельного алгоритма анализа данных, позволяющая проводить распараллеливание как по данным, так и по задачам для выполнения в общей и распределенной памяти;
- метод распараллеливания последовательных алгоритмов машинного обучения как по задачам, так и по данным, формирующий оптимальное число параллельных ветвей, с учетом особенностей их выполнения;
- метод размещения функций алгоритмов анализа данных в среде выполнения, минимизирующий время выполнения алгоритма и сетевой трафик с учетом способа хранения данных;
- методика построения параллельных алгоритмов анализа данных, использующая предложенные модели и методы и позволяющая оптимизировать структуру параллельного алгоритма для минимизации времени выполнения и сетевого трафика с учетом условий выполнения.

Результаты исследований представлены в более чем 20 публикациях, в том числе ведущих журналах в области компьютерных наук. По результатам были защищены одна докторская диссертация и две диссертации кандидата наук.

Сотрудники принимали участие в проекте Седьмой рамочной программы (FP7) Европейского Сообщества, в рамках которого разрабатывалась система управления информационной безопасностью нового поколения, ею разрабатывались проблемы обнаружения различных сетевых атак и своевременного принятия контрмер, построения системы визуального анализа инцидентов безопасности, а также вопросы, связанные с проектированием архитектуры системы нового поколения (Управление информацией и событиями безопасности в инфраструктурах услуг (MASSIF)). Участвовали в выполнении ряда проектов, поддержанных российскими научными фондами (РНФ и РФФИ).

Для практического применения, полученные результаты были реализованы в виде библиотеки алгоритмов машинного обучения для анализа распределенных данных (FL4J <https://gitlab.fkti.etu.ru/fl4j>). Библиотека реализована на языке программирования Java, что позволяет ее использовать в промышленных разработках.

В библиотеке алгоритмы реализованы в виде отдельных функций алгоритма, что соответствует функциональной модели алгоритма.

В результате библиотека, имеет следующие возможности по сравнению с существующими библиотеками алгоритмов анализа данных:

- построение новых алгоритмов путем комбинации функций или модификация существующих алгоритмов путем замены (или модификации) отдельных функций;
- преобразование последовательных алгоритмов, построенных в виде последовательности функций, в параллельные формы для выполнения в параллельных и распределенных средах;
- выполнение параллельных алгоритмов анализа данных в разных вычислительных средах.

Библиотека содержит необходимые базовые классы для построения параллельных алгоритмов анализа данных. Она разделена на следующие модули:

- ядро библиотеки (Core), построенное в соответствии со стандартом CWM и включающее в себя:
 - средства работы с данными;
 - средства сохранения результатов, в соответствии с предложенным представлением модели знаний;
 - средства для настройки задач и алгоритмов анализа данных;

- базовые классы для реализации ФОМ, в соответствии с функциональной моделью алгоритма анализа данных;
- программную реализацию исполнителей для выполнения в параллельной и распределенной среде;
- алгоритмы классификации (Classification);
- алгоритмы кластеризации (Clustering);
- алгоритмы поиска ассоциативных правил (Association).

Последние три модуля используют ядро и не зависят друг от друга (см. Рисунок 2).

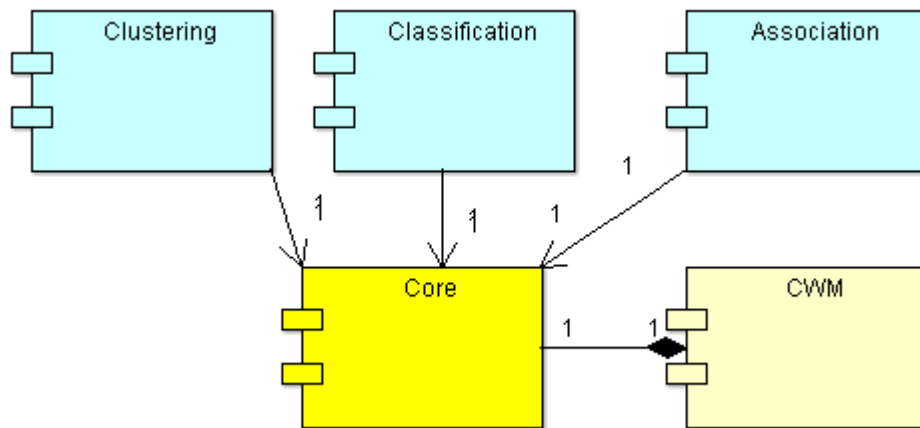


Рис. 2 – Структура компонентов библиотеки.

Членами научного коллектива в рамках работ компании ООО “Смартлайзер Рус” выполнены следующие работы по аналогичной тематике:

1. Разработана система планирования и отслеживания работ по утилизации строительных отходов на заводах компании FORTUM, подразделение Fortum Waste Solutions. Система внедрена на 3-х заводах по переработке строительных отходов в городах Norrtoorp, Sundsvall и Kiruna. Разработанная система является расширением корпоративной ERP. Примеры пользовательского интерфейса менеджера и оператора (водителя погрузчика, самосвала, экскаватора и т.п.) приведены на рисунках 3, 4, и 5.

Указанная система реализует расширенный набор функций, например, поддерживает режимы эвакуации и Inactivity alerting (предупреждение об отсутствии активности оператора), обеспечивает логистическое сопровождение доставки строительных отходов на завод.

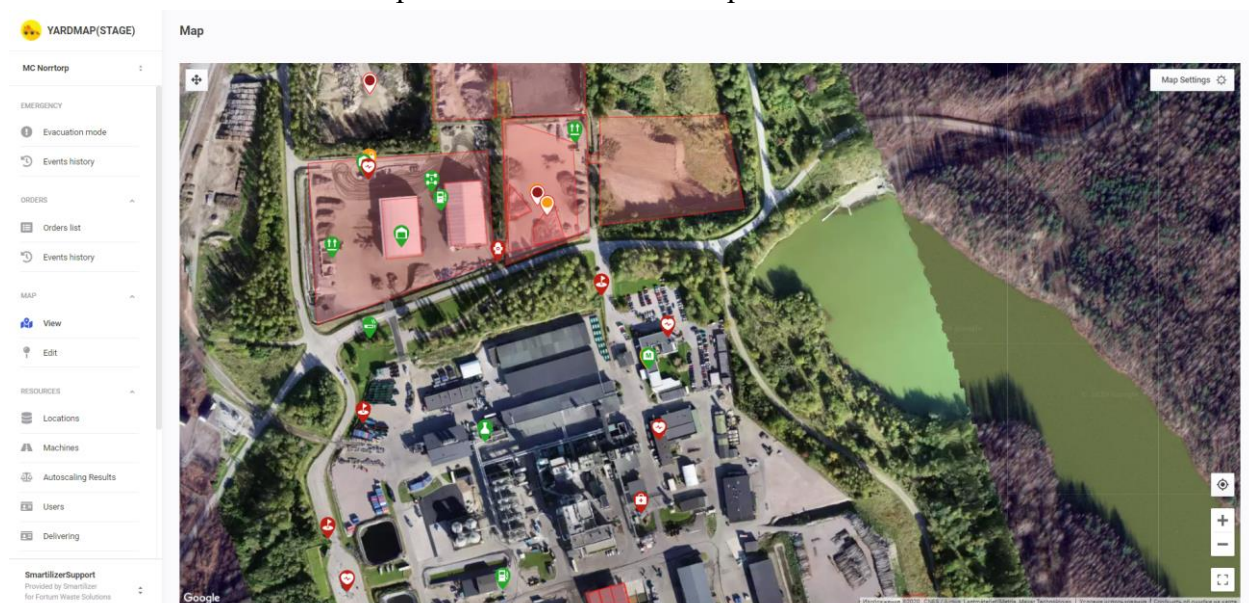


Рис.3. Пример интерфейса менеджера (гео-позиционирование устройств и объектов)

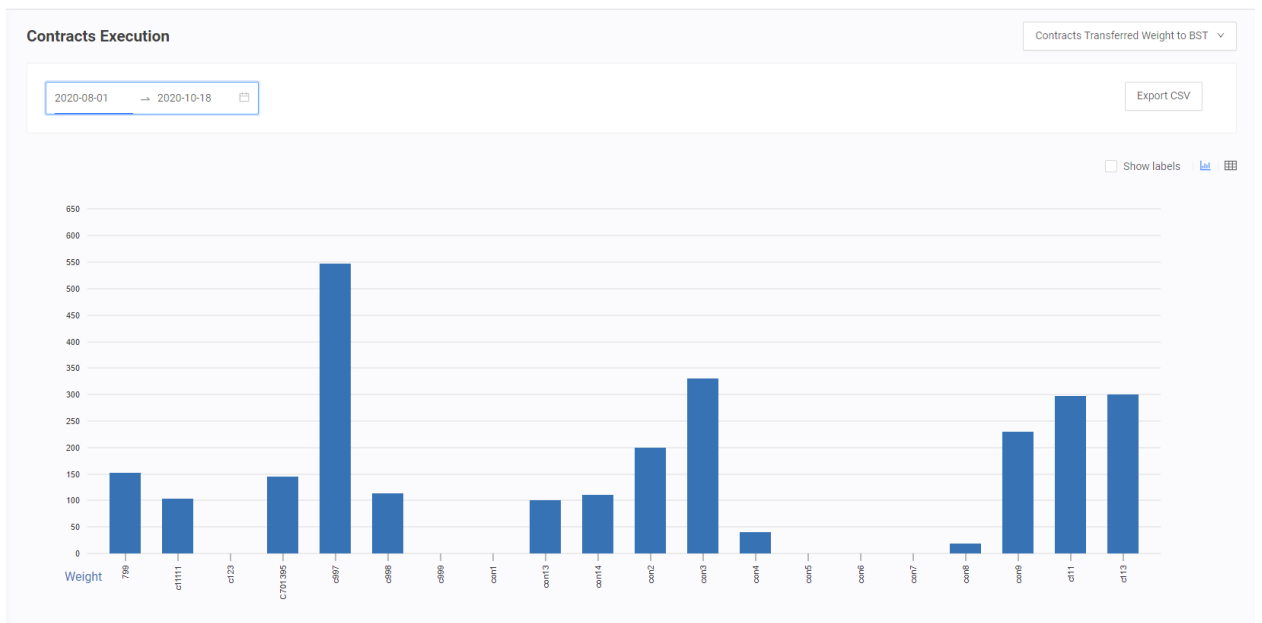


Рис.4. Пример интерфейса менеджера (метрики)

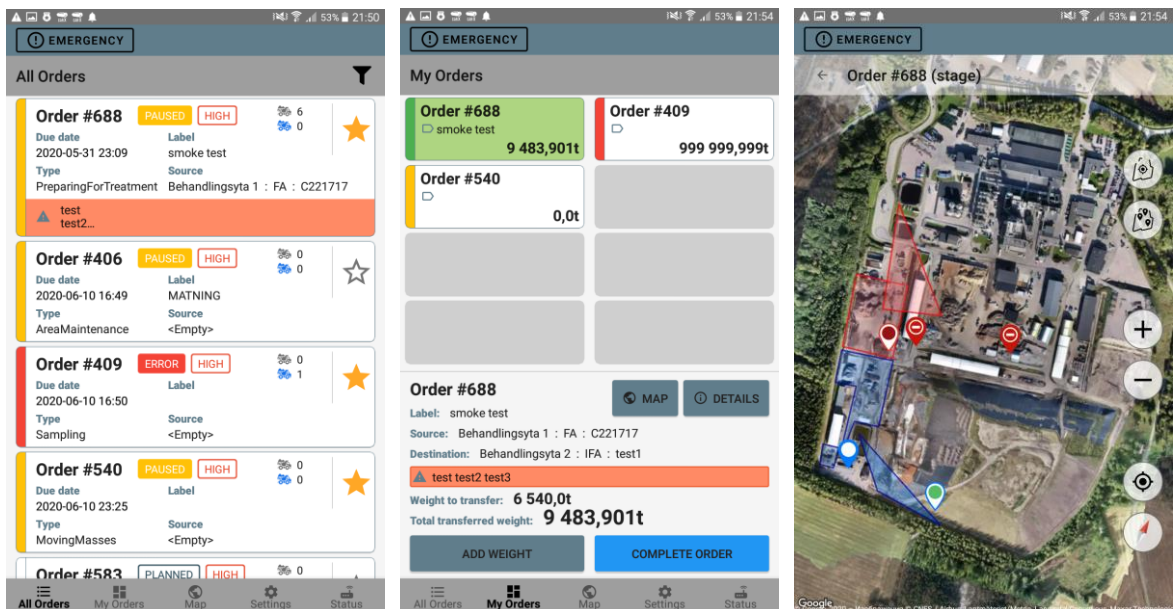


Рис.5. Примеры интерфейсов оператора

2. Разработано приложение для работы с 3D-моделями зданий и сооружений при проектировании, строительстве и дальнейшем обслуживании указанных сооружений. Примеры отображения 3D-модели в веб-приложении и на мобильном устройстве приведены на рисунке 6 и 7.

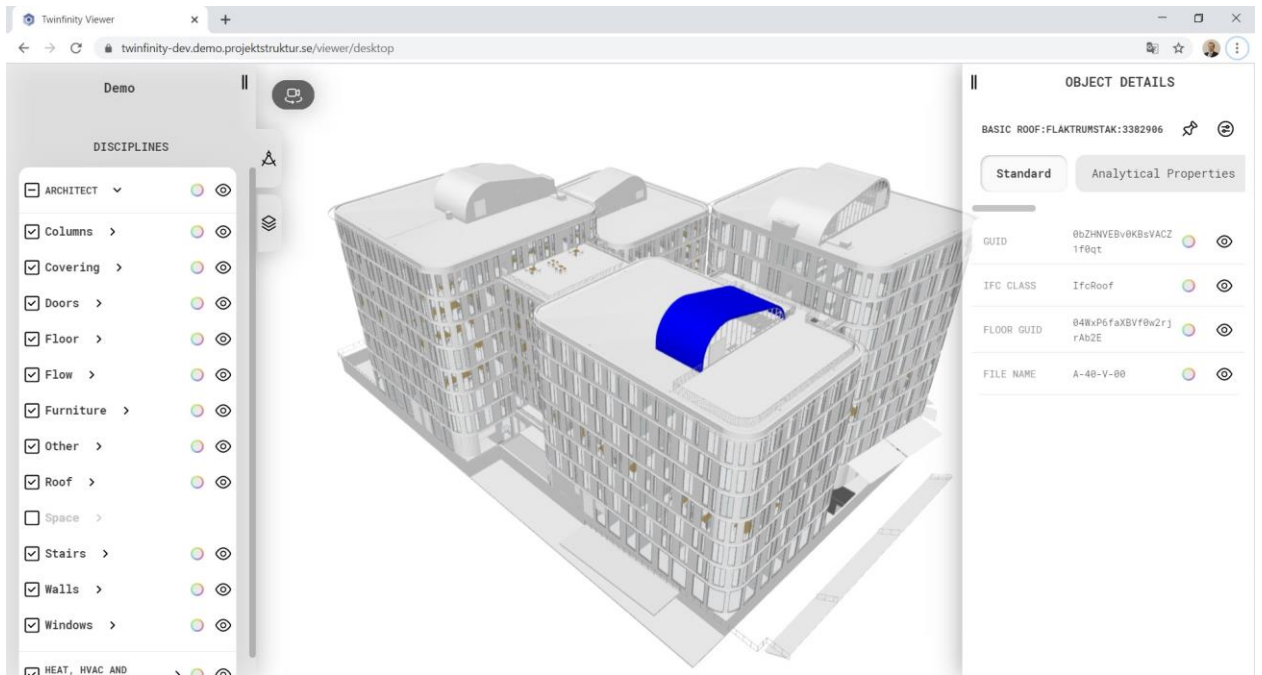


Рис.6. Пример отображения 3d-модели в веб-приложении

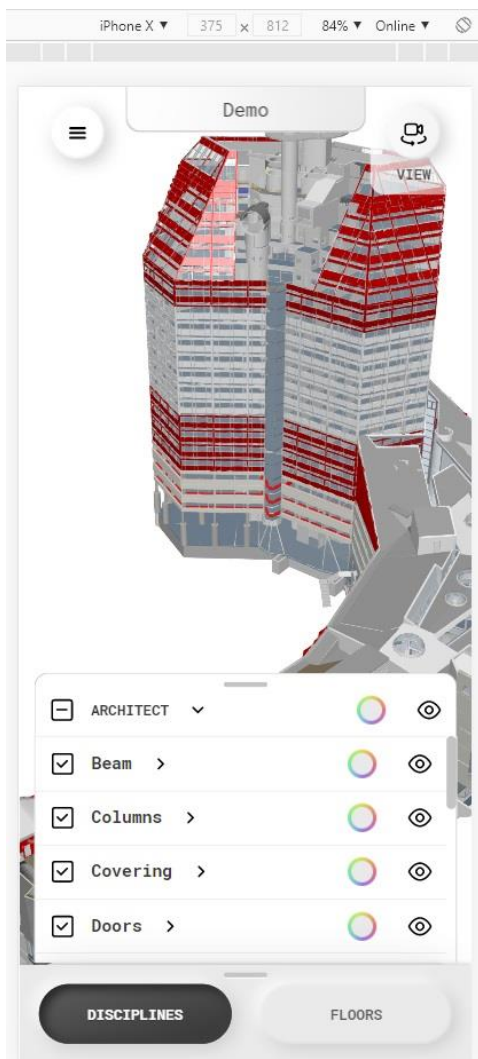


Рис.7. Пример отображения 3d-модели в мобильном приложении

5.3. Организационная структура управления.



Рис.8. Организационная структура проекта

6. ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ РЕАЛИЗАЦИИ ПРОЕКТА

6.1. Календарный план реализации проекта

Все даты в календарном плане поставлены исходя из предположения, что работа начинается 9 января 2023 года.

Проект реализуется в четыре этапа.

Первый этап посвящен сбору требований, разработке архитектуры системы, подготовке расширенного технического задания, и прототипированию. На этом этапе разрабатывается прототип, в котором вся информация с датчиков в течение рабочего дня накапливается на мобильных устройствах и пересылается в облачный сервис по беспроводному каналу. В облачном сервисе осуществляется анализ данных на основе нейронной сети, обученной по методу «обучение с учителем». На этом этапе проводится исследование доступных сред разработки с открытым исходным кодом и выбор наиболее подходящей среды разработки для использования на последующих этапах.

Второй этап посвящен разработке технологии распределенной обработки данных на базе федеративного обучения (FL) и визуализации полученных результатов в виде отчетов и метрик. На этом этапе планируется реализация распределенных вычислений с перемещением нейронной сети на мобильные устройства с реализацией федеративного обучения нейронной сети (FL – Federated Learning), что позволит обрабатывать данные о работе машин в оперативном режиме и локально доучивать модель искусственного интеллекта в процессе работы машин. Что будет позволять избежать сложного для водителей и организаций этапа «доучивания» нейронной сети при добавлении новых машин в рабочий процесс при централизованной организации обработки данных с помощью искусственного интеллекта.

Третий этап посвящен интеграции платформы с ERP системой заказчика и разработке технологии прогнозирования необходимого количества машин на основе статистических данных и дополнительных факторов. На этом этапе планируется использование результатов, полученных от распределенной нейронной сети (Federated Learning), данных о планируемых работах, полученных от ERP заказчика, и ряда иных общедоступных данных, тип которых зависит от поставленных заказчиками задач (финансовые прогнозы, погода, обстановка на

дорогах и т.п.) для прогнозирования необходимого количества машин, внедрение разработанной системы как минимум на одном из предприятий заказчиков.

Этот этап также направлен на коммерциализацию разрабатываемого решения: доработка системы по результатам пробной эксплуатации, расширение возможностей облачного сервиса, поддержка системы, активная реклама, подготовка и заключение договоров.

Таблица 4. Календарный план проекта с указанием поставок по этапам

№	Наименование этапа	Ключевые контрольные точки ¹	Плановый срок	Объем финансирования, тыс.руб.
1	Сбор требований, разработка архитектуры системы, подготовка расширенного технического задания, прототипирование	Прототип вер.0.1 для работы с типами данных заказчика	28.02.2023	10 000
		Подписание соглашений для установки клиентских приложений на стороне владельцев данных	28.02.2023	
		Разработка моделей нейронной сети вер.0.1	31.03.2023	
		Разработка моделей нейронной сети федеративного типа вер.0.2	31.03.2023	
		Формирование наборов данных для обучения нейронных сетей вер. 0.1 и 0.2	28.03.2023	
		Проведение испытаний с нейронными сетями установленными на сервере и на клиентах	31.04.2023 30.05. 2023	
		Подготовка расширенного технического задания	30.06. 2023	

		Разработка архитектуры системы	30.06. 2023	
		Подготовка статьи и выступления на конференции	30.06. 2023	
		Отчет по этапу 1	30.06. 2023	
2	Разработка технологии распределенной обработки данных на базе федеративного обучения (FL), визуализация полученных результатов в виде отчетов и метрик	Разработка моделей распределенной обработки данных вер. 0.3	31.08. 2023	10 000
		Выбор механизмов и алгоритмов защиты данных в среде разработки федеративного обучения	31.08. 2023	
		Реализация алгоритмов защиты данных в выбранной среде разработки федеративного обучения с открытым исходным кодом	31.10. 2023	
		Проведение экспериментов на подготовленных на этапе 1 наборах данных	31.10. 2023	
		Разработка модуля формирования отчетов и визуализации метрик	31.10. 2023	
		Подготовка статьи и выступления на конференции	31.10. 2023	
		Оформление программ формирования наборов данных и аннотирования данных в ФИПС	31.12. 2023	

		Подписание соглашений на проведение испытаний системы в реальных условиях	31.10. 2023	
		Проведение испытаний в полевых условиях	31.12. 2023	
		Проведение рекламной кампании	на протяжении этапа	
		Отчет по этапу 2	31.12. 2023	
3	Интегрирование с ИС/ERP системой заказчика и разработка технологии прогнозирования необходимого количества машин на основе статистических данных и дополнительных факторов	Разработка модели прогнозирования	28.02.2024	10 000
		Разработка модуля интеграции с ERP системой заказчика	28.02.2024	
		Проведение рекламной кампании	на протяжении этапа	
		Настройка параметров алгоритмов анонимизации данных и дифференциальной безопасности данных, используемых в моделях прогнозирования	30.04.2024	
		Подготовка статьи и выступления на конференции. Оформление программных моделей распределенной обработки данных и прогнозирования в ФИПС	30.04. 2024	

		Проведение испытаний в полевых условиях	30.06. 2024	
		Расширение возможностей облачного сервиса	30.06. 2024	
		Отчет по этапу 3	30.06. 2024	
4	1-5-й годы после НИОКР. Коммерциализация: расширение возможностей облачного сервиса, поддержка системы, реклама, заключение договоров; внедрение системы на многих предприятиях	Поддержка системы		Увеличивается от 5 000 тыс. руб. до 8 000 тыс. руб. в квартал
		Проведение рекламной кампании		
		Подготовка договоров, лицензионных соглашений		
		Заключение договоров с заказчиком		
		Развитие нового функционала по результатам внедрения		
		Рекламная кампания		

6.2. Наличие основных средств и необходимых площадей для реализации проекта, в том числе для выполнения НИОКР.

Для реализации проекта будет использован арендованный офис, находящийся по адресу: Санкт-Петербург, наб. Обводного канала, д.93А, бизнес-центр Нептун, офис 12.

7. ФИНАНСОВЫЙ ПЛАН

Для реализации проекта потребуются инвестиции в НИОКР на протяжении 1,5-х лет в объеме примерно 30 млн. руб.

После завершения НИОКР, проектной команде потребуется порядка 5 000 тыс. руб. в квартал на поддержание инфраструктуры, сопровождение разработанного и доработку нового функционала, рекламную кампанию, разработку PoC и MVP проектов для новых заказчиков.

Подход с использованием сбалансированной по опыту и навыкам работы команды позволит обеспечить стоимость команды из 4-5 человек порядка 600 тыс. руб. в месяц и выполнять проекты типа PoC за 2 месяца при стоимости 1,2 млн. руб., выполнять проекты типа MVP за 8 месяцев при стоимости порядка 7,2 млн. руб. и обеспечивать сопровождение внедренных решений при стоимости 250 тыс. руб. в месяц.

Предполагаемые затраты, выручка и рентабельность показаны на рисунках 9, 10 и 11

Платформа анализа данных - затраты																								
	2023				2024				2025				2026				2027				2028			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Затраты на этапе НИОКР (24 месяца), тыс. руб	5 000	5 000	5 000	5 000	5 000	5 000																		
Затраты на поддержание инфраструктуры, тыс. руб							200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200	200
Затраты на коммерциализацию, тыс. руб							4 800	4800	4800	4800	4800	4800	5 800	5800	5800	5800	6800	6800	6800	6800	7800	7800	7800	7800
ИТОГО по годам, тыс.руб.				20 000			20 000				20000				24000				28000				32000	
Себестоимость за 5 лет, тыс. руб																							144000	

Рис.9. Затраты, себестоимость и варианты продаж с линейным и мультипликативным ростом

Платформа анализа данных - выручка															
Активности	Стоимость 1 мес., тыс.руб	Длительность, месяц	2023		2024		2025		2026		2027		2028		ВСЕГО
			шт.	тыс.руб	шт.	тыс.руб	шт.	тыс.руб	шт.	тыс.руб	шт.	тыс.руб	шт.	тыс.руб	
РоС проекты (4-5 чел.)	600	2	1	1200	3	3600	6	7200	9	10800	12	14400	15	18000	55200
MVP проекты (4-5 чел.)	900	8	0	0	1	7200	2	14400	3	21600	4	28800	5	36000	108000
Поддержка развернутых решений	250	12	0	0	0	0	1	3000	2	6000	3	9000	4	12000	30000
ВЫРУЧКА				1200		10800		24600		38400		52200		66000	193200

Рис.10. Затраты, себестоимость и варианты продаж с линейным и мультипликативным ростом

Платформа анализа данных - рентабельность по годам							
	2023	2024	2025	2026	2027	2028	ВСЕГО
Себестоимость производства, тыс. руб	20 000	20000	20000	24000	28000	32000	144 000
Выручка за 5 лет, тыс. руб	1 200	10 800	24 600	38 400	52 200	66 000	193 200
Прибыль за 5 лет, тыс. руб	-18 800	-9 200	4 600	14 400	24 200	34 000	49 200
Налог 3% (для аккредитованной ИТ компании), тыс. руб	0	0	138	432	726	1 020	1 476
Чистая прибыль, тыс. руб	-18 800	-9 200	4 462	13 968	23 474	32 980	47 724
Рентабельность	-94,0%	-46,0%	22,3%	58,2%	83,8%	103,1%	33,1%

Рис.11. Рентабельность 5 лет при линейном и мультипликативном росте продаж

7.5. Возможные моменты, типы и источники рисков, меры по их уменьшению.

Таблица 9. Проектные риски и планы по их уменьшению

No.	Название риска / Краткое описание	Вероятность (от 0 до 9)	Влияние (low, medium, high)	Ранг	План по устранению или сглаживанию
1	Решение может оказаться менее интересным для заказчика чем у конкурентов	4	M	M	1. Работа не с одним, а с несколькими заказчиками 2. Регулярное обсуждение функционала с потенциальными заказчиками, доработка требований 3. Анализ рынка регулярно
2	Объем бизнеса заказчика может оказаться ниже запланированного	5	L	L	1. Работа не с одним, а с несколькими заказчиками 2. Доработка требований для расширения возможностей системы и более широкого внедрения 3. Поиск новых заказчиков

3	Заказчик может изменить требования к продукту, что потребует доработки ПО	5	M	M	<ol style="list-style-type: none"> 1. Использование гибких (agile) моделей разработки 2. Регулярное обсуждение функционала с потенциальными заказчиками, доработка требований 3. Регулярные демонстрации продукта
4	Риски, связанные с коммуникацией между разработчиком и заказчиком	3	L	L	<ol style="list-style-type: none"> 1. Регулярная проектная отчетность 2. Регулярные демонстрации готовой функциональности заказчикам 3. Проведение совместных митингов по планированию и ретроспективе с заказчиками
5	Недостаточно детализированная документация по проекту	4	L	L	<ol style="list-style-type: none"> 1. Использование современных средств документирования (Wiki) 2. Использование канбан-доски и системы отслеживания задач (Jira) 3. Подготовка/обновление демонстраций продукта после каждого спринта
6	Недооценка требуемых усилий	4	M	M	<ol style="list-style-type: none"> 1. Использование гибких (agile) моделей разработки 2. Проведение предварительной (по требованиям в Wiki) и детальной (по задачам в Jira) оценок 3. Определение производительности команды (velocity) в начале разработки на первых спринтах
7	Реализация мало используемой функциональности	4	M	M	<ol style="list-style-type: none"> 1. Использование гибких (agile) моделей разработки 2. Приоритезация функционала на уровне требований в wiki и на уровне задач в Jira 3. Регулярное обновление приоритетов по результатам совместных совещаний/демонстраций
8	Ошибка бюджетирования	3	M	M	<ol style="list-style-type: none"> 1. Проведение операционных обзоров и отслеживание финансовых показателей ежемесячно 2. Представление финансовых результатов совету директоров ежеквартально 3. Формирование детальных отчетов по гранту за каждый этап
9	Несвоевременное финансирование проекта	4	H	H	<ol style="list-style-type: none"> 1. Использовать свободные средства от других проектов 2. Использовать займы от владельцев компании
10	Кадровые риски (выход из рабочего процесса разработчика ПО, тестировщика ПО)	4	M	M	<ol style="list-style-type: none"> 1. Формировать команду из универсальных разработчиков, способных заменять друг друга 2. Поддерживать университетскую программу по подготовке кадров

					3. Использовать агентства для быстрого найма
11	Уровень квалификации участников проекта не соответствует поставленным задачам	5	М	М	<ol style="list-style-type: none"> 1. Формировать сбалансированную команду (архитектор, старшие инженеры, инженеры) 2. Начинать прототипирование на ранней стадии 3. Разрабатывать архитектуру системы и ее детализацию с привлечением всей команды
12	Недостаточная производительность системы	4	М	М	<ol style="list-style-type: none"> 1. Описать нефункциональные требования 2. Учитывать их при разработке архитектуры 3. Подготовить стресс тесты и тесты производительности

Правила работы с рисками, оценивания вероятности и влияния, формирование ранга риска должны быть описаны в стандартном процессе работы компании и доведены до всех участников проекта.

Учредитель ООО «Смартилайзер Рус»
 Филиппов Евгений Васильевич
 19 декабря 2022г.



Приложение 1. Литература

1. H Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. “Communication-Efficient Learning of Deep Networks from Decentralized Data”. <https://arxiv.org/abs/1602.05629>
2. Kairouz, Peter et al. “Advances and Open Problems in Federated Learning.” ArXiv abs/1912.04977 (2019): n. pag.
3. Rieke, N., Hancox, J., Li, W. et al. The future of digital health with federated learning. *npj Digit. Med.* 3, 119 (14 Sept 2020). <https://doi.org/10.1038/s41746-020-00323-1>
4. P. Voigt and A. Von dem Bussche. The EU general data protection regulation (GDPR). A Practical Guide, 1st Ed., Cham: Springer International Publishing, 2017.
5. California Consumer Privacy Act Home Page. <https://www.caprivacy.org/>.
6. 2019 is a 'fine' year: Pdpcc has fined s'pore firms a record \$1.29m for data breaches, 2019. URL <https://vulcanpost.com/676006/pdpcc-data-breach-singapore-2019/>.
7. Cybersecurity Law of the People's Republic of China (English Translation), effective June 1, 2017: <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-cybersecurity-law-peoples-republic-china/>
8. Instruction of Paddle FL: <https://paddlefl.readthedocs.io/en/latest/index.html>
9. McMahan, H. Brendan et al. “Communication-Efficient Learning of Deep Networks from Decentralized Data.” In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, pages 1273–1282, 2017 (original version on arxiv Feb. 2016).
10. Keith Bonawitz and Vladimir Ivanov and Ben Kreuter and Antonio Marcedone and H. Brendan McMahan and Sarvar Patel and Daniel Ramage and Aaron Segal and Karn Seth. Practical Secure Aggregation for Federated Learning on User-Held Data. <https://arxiv.org/abs/1611.04482>
11. Ali Davody and David Ifeoluwa Adelani and Thomas Kleinbauer and Dietrich Klakow. Robust Differentially Private Training of Deep Neural Networks. <https://arxiv.org/abs/2006.10919>
12. Abadi, Martin et al. “Deep Learning with Differential Privacy.” Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (2016): n. pag. Crossref. Web. <https://arxiv.org/abs/1607.00133>
13. Renuga Kanagavelu and Zengxiang Li and Juniarto Samsudin and Yechao Yang and Feng Yang and Rick Siow Mong Goh and Mervyn Cheah and Praewpiraya Wiwatphonthana and Khajonpong Akkarajitsakul and Shangguang Wangz. Two-Phase Multi-Party Computation Enabled Privacy-Preserving Federated Learning. <https://arxiv.org/abs/2005.11901>
14. Payman Mohassel and Peter Rindal. ABY3 : A Mixed Protocol Framework for Machine Learning. <https://eprint.iacr.org/2018/403.pdf>
15. https://github.com/PaddlePaddle/PaddleFL/tree/master/python/paddle_fl/mpc/examples/model_encryption
16. Wikipedia. Technology readiness level. https://en.wikipedia.org/wiki/Technology_readiness_level
17. <https://realty.rbc.ru/news/5e8341eb9a794755f5f2ce96>
18. https://infraone.ru/sites/default/files/analitika/2020/infraone_research_weekly_13_37_30062020.pdf
19. https://aebrus.ru/upload/iblock/33e/rus_-construction-equipment-for-q-4-2019.pdf
20. <https://www.garant.ru/products/ipo/prime/doc/72175880/>
21. <https://rg.ru/2020/08/06/rynok-dorozhno-stroitelnoj-tehniki-v-rossii-podderzhat-krupnye-proekty.html>
22. <https://www.grandviewresearch.com/industry-analysis/construction-equipment-rental-market>

23. Novikova, E.; Fomichov, D.; Kholod, I.; Filippov, E. Analysis of Privacy-Enhancing Technologies in Open-Source Federated Learning Frameworks for Driver Activity Recognition. *Sensors* 2022, 22, 2983. <https://doi.org/10.3390/s22082983>
24. Efremov, M.A., Kholod, I.I., Kolpaschikov, M.A.: Java federated learning framework C1-230914: 03.09.2022 Страница 13 из 33 architecture. In: 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), pp. 306–309 (2021). DOI 10.1109/ElConRus51938.2021.9396508
25. Kholod, I.; Yanaki, E.; Fomichev, D.; Shalugin, E.; Novikova, E.; Filippov, E.; Nordlund, M. OpenSource Federated Learning Frameworks for IoT: A Comparative Review and Analysis. *Sensors* 2021, 21, 167. <https://doi.org/10.3390/s21010167>
26. Ivan Kholod, Andrey Shorov, and Sergei Gorlatch. Efficient Distribution and Processing of Data for Parallelizing Data Mining in Mobile Clouds. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, 11(1):2-17, Mar. 2020 DOI:10.22667/JOWUA.2020.03.31.002
27. Kholod, I., Rukavitsyn, A., Paznikov, A. et al. Parallelization of the self-organized maps algorithm for federated learning on distributed sources. *J Supercomput* (2020). <https://doi.org/10.1007/s11227-020-03509-2>
28. M. A. Efremov and I. I. Kholod, "Architecture of Swarm Robotics System Software Infrastructure," 2020 9th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 2020, pp. 1-4, doi: 10.1109/MECO49872.2020.9134247.
29. Ivan Kholod, Andrey Shorov, and Sergei Gorlatch. Improving Data Mining for Differently Distributed Data in IoT Systems // The 13th International Symposium on Intelligent Distributed Computing (IDC 2019) pp 75-85 DOI: 10.1007/978-3-030-32258-8_9
30. Kholod I. I. et al. FL4J—Federated Learning Framework for Java //Intelligent Distributed Computing XIV. – С. 225. DOI: 10.1007/978-3-030-96627-0_21
31. Холод И.И., Малов А.В., Родионов С.В. Способ распараллеливания интеллектуального анализа данных в вычислительной среде. //Патент на изобретение №2745018, 18 марта 2021 г.
32. Ефремов М.А., Колпашиков М.А., Табаков П.Л. Программный адаптер для управления сервером федеративного обучения. //Свид. о государств. Регистрации программы для ЭВМ № 2021669639 от 23.11.2021.
33. Ефремов М.А., Табаков П.Л. Программа для регистрации клиентов федеративного обучения. //Свид. о государств. Регистрации программы для ЭВМ № 2021669370 от 23.11.2021.
34. Ефремов М.А., Аристархов И.Е. Программа управления клиентом федеративного обучения. //Свид. о государств. Регистрации программы для ЭВМ № 2021669391 от 23.11.2021.
35. Холод И.И. Программа подготовки набора функциональных блоков интеллектуального анализа данных к параллельному выполнению.//Свид. о государств. Регистрации программы для ЭВМ № 2016610772 от 19.01.2016