

# **DS** **DATASKAI**

Фреймворк для AI/ML-проектов

(R&D-версия)

Описание процессов разработки и поддержки продукта

**Skoltech**

Веб-страница проекта: <http://dataskai.com>

## 1 Состав пакета и способ доступа

Фреймворк DATASKAI доступен как в полной версии, размещенной в сети «Интернет» на сайте <http://dataskai.com>, так и в виде пакета изолированной установки.

Пакет изолированной установки поставляется по запросу.

## 2 Класс программного обеспечения

Фреймворк относится к классу «Средства обработки Больших Данных (BigData)» раздела «Средства обработки и визуализации массивов данных» Классификатора программ для электронных вычислительных машин и баз данных, утвержденного приказом Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации от 22 сентября 2020 г. № 486.

Код Общероссийского классификатора продукции по видам экономической деятельности: 62 63.11.19.

Код класса программ для электронных вычислительных машин и баз данных: 10.01.

## 3 Процессы разработки и поддержки программного продукта

### 3.1 Стадии и этапы разработки

Разработка проводится по следующим этапам, указанным в таблице 1.

Разработка, кастомизация и поддержка продукта выполняются специалистами ЦК НТИ БСИВ на базе Сколковского института науки и технологий.

Процесс поддержки продукта предусматривает организацию двух линий поддержки на базе ЦК НТИ БСИВ в 2022 г., доступных как по электронной почте, так и по контактными телефонам, размещенным на интернет-сайте <http://iot.skoltech.ru>.

В ходе процесса разработки были установлены и уточнены требования к функциональным характеристикам, функции продукта, характеристики пользователей, а также ограничения, допустимые при разработке.

## 3.2 Требования к программе или программному изделию

### 3.2.1 Требования к функциональным характеристикам

Платформа должна осуществлять сбор данных путем взаимодействия с источниками данных: хранилищами, информационными системами и устройствами.

Примерами хранилищ являются файловые серверы, СУБД, очереди, распределенные хранилища и файловые системы.

Примерами информационных систем служат различного рода корпоративные системы (ERP, SCADA, CSM, CRM, SRM, MRP и т.п.), обладающие интерфейсами для выгрузки данных.

Примерами устройств являются устройства, генерирующие потоки данных (всевозможные датчики, устройства интернета вещей и т.п.).

Взаимодействие платформы с источниками должно производиться с использованием поддерживаемых протоколов, либо, при отсутствии поддержки со стороны источника, через отдельно разрабатываемые интеграционные модули.

Платформа должна предоставлять доступ внешним приложениям к собранным данным и результатам их анализа через открытые программные интерфейсы.

### 3.2.2 Функции продукта

Платформа должна выполнять следующие основные функции:

- сбор данных из источников данных и их преобразование во внутренние представления для эффективного хранения, обработки и анализа;
- хранение внутренних представлений данных и реализация унифицированного доступа к данным через внешние интерфейсы;
- обработку данных с возможностью встраивания экспертных моделей предметной области;
- анализ с возможностью формализации постановок и контроля решения различных классов задач на данных.

### 3.2.3 Характеристики пользователя

Пользователями платформы должны являться:

- специалист по обработке данных, решающий задачи анализа данных с использованием платформы;

- разработчик внешних приложений, реализующий приложения для конечных потребителей результатов решения задач анализа данных;
- системный администратор, осуществляющий развертывание и сопровождение платформы;
- инженер по данным, решающий задачи по организации сбора и хранения данных с использованием платформы.

### 3.2.4 Ограничения

С учетом специфики применения платформы ее реализация имеет следующие характерные особенности:

- ориентация на работу с данными, зависящими от времени, а также табличными данными, связанными реляционными отношениями;
- не предполагается работа с данными, хранимыми в машиночитаемом графическом виде, однако с заложенной возможностью расширения на другие типы данных с использованием открытого API;
- не предполагается работа с данными в режиме реального времени, вместо этого платформа нацелена на работу с данными в потоковом режиме;
- поток входящих данных для сбора и хранения не более 1ТБ / сутки.

### 3.2.5 Требования к интерфейсам пользователя

Платформа должна содержать веб-интерфейс для специалиста работы с данными, в котором будут отображены оценки качества работы алгоритмов на данных.

Платформа должна содержать веб-интерфейс системного программиста для отслеживания состояния компонентов системы, их журналов для предупреждения сбоев и отказов.

Платформа должна содержать интерфейс системного администратора для базовой настройки компонентов при развертывании и эксплуатации Платформы.

Платформа должна содержать веб-интерфейс инженера по данным для отслеживания процессов сбора и обработки данных.

### 3.3 Сведения о команде разработчиков

В состав команды разработчиков (12 чел. по состоянию на октябрь 2021 г.) входят сотрудники, выполняющие следующие роли:

- руководитель проекта,
- главный разработчик программного обеспечения,
- ведущий аналитик,
- инженер по внедрению,
- инженеры-программисты (3 чел.),
- инженеры-исследователи (2 чел.),
- проектировщик графических интерфейсов,
- специалист службы поддержки (2 чел.).

#### **Главный разработчик программного обеспечения**

- Разрабатывает алгоритмы для автоматической обработки массивов данных на основе существующих методов анализа данных и экспертизы специалистов из предметной области со стороны заказчиков;

- Реализует алгоритмы в виде прототипов ПО и проводит отладку их работы;

- Тестирует алгоритмы в соответствии со стандартной методологией анализа данных;

- Разрабатывает рабочие планы программы выполнения отдельных этапов работ, методики исследований, испытаний, математической и графической обработки результатов измерений;

- Обрабатывает, анализирует и обобщает результаты исследований, составляет отчеты по теме или ее части, согласовывает разрабатываемые материалы с непосредственным руководителем и с представителями партнеров;

- Участвует во внедрении результатов исследований и разработок, оказании экспертной помощи при проектировании, изготовлении, испытаниях и сдаче ПО;

- Составляет описание программ и алгоритмов, инструкции по эксплуатации и инструкции на проведение исследовательских работ;

- Участвует во взаимодействии с заказчиками на предмет разработки технических решений;

- Осуществляет анализ и разработку требований к продуктам;

- Осуществляет руководство группой разработчиков ПО, определяет задачи для разработчиков, стажеров-исследователей, инженеров и иных работников, в отношении которых Работник осуществляет функциональное руководство;

- Отвечает за техническую реализацию порученных проектов.

### **Инженер-программист**

- Участвует в развитии программной архитектуры компонентов платформы;
- Выполняет перепроектирование и переработку компонентов платформы;
- Участвует в работе над формализацией задач машинного обучения с использованием компонентов платформы;
- Участвует в контроле процесса разработки ПО, проводит обзор кода (code review) и консультирует по техническим вопросам других разработчиков;
- Выполняет разведочный анализ данных;
- Выполняет решение формализованных задач машинного обучения и оптимизации с использованием (без использования) специализированных инструментов DATASKAI;
- Содействует развитию и улучшению рабочего процесса и инструментов работы с использованием фреймворка в рамках своих компетенций;
- Участвует в решении смежных задач ЦК НТИ в рамках своих компетенций;
- Создает, отлаживает и оптимизирует программный код;
- Составляет автоматизированные тесты;
- Содействует решению проблем при эксплуатации ПО в операционном окружении в рамках своих компетенций;
- Составляет техническую документацию к разработанному ПО;
- Участвует в разработке архитектуры компонентов;
- Разрабатывает модели данных в рамках проектирования реляционных и не реляционных баз данных;
- Разрабатывает сценарии непрерывной интеграции и доставки программного обеспечения (CI/CD);
- Содействует решению инфраструктурных проблем в работе программного обеспечения на серверах ЦК НТИ в рамках своих компетенций.

### **Инженер-исследователь**

- Выполняет решение формализованных задач машинного обучения и оптимизации с использованием (без использования) специализированных инструментов;
- На основе существующих методов анализа данных и экспертизы специалистов из предметной области разрабатывает алгоритмические цепочки для автоматической обработки массивов данных;
- Тестирует алгоритмические цепочки в соответствии со стандартной методологией анализа данных;

- Создает, отлаживает и оптимизирует программный код;
- Выполняет проектирование и разработку веб-сервисов;
- Выполняет проектирование архитектуры баз данных;
- Участвует во внедрении результатов исследований и разработок, осуществлении авторского надзора при проектировании, изготовлении, монтаже, наладке, испытаниях и сдаче опытных образцов изделий в эксплуатацию;
- Выполняет разработку алгоритмов непрерывной интеграции и доставки кода (CI/CD);
- Составляет техническую документацию к разработанному ПО;
- Участвует в подготовке отчетов по НИР, подготовке результатов интеллектуальной деятельности к оформлению;
- Участвует в планировании задач по разработке программного обеспечения.

### **Проектировщик графических интерфейсов**

- Работает над улучшением UI/Usability/UX-качеств продуктов;
- Разрабатывает и поддерживает в актуальном состоянии руководства по стилю интерфейса (UI Guidelines);
- Разрабатывает сценарии тестирования интерфейсов;
- Тестирует на удобство использования (usability testing) и определение влияния эстетики на удобство использования (usability);
- Проектирует внутренние графические интерфейсы для конфигурирования/администрирования собственных сервисов/инструментов;
- Проектирует внешние, проблемно-ориентированные веб-интерфейсы для систем предсказательного моделирования и анализа данных;
- Работает с требованиями UI/UX.

### **Инженер по внедрению**

- Разрабатывает CI/CD процессы;
- Разрабатывает Ansible плейбуки;
- Создает распределенные и отказоустойчивые сервисы;
- Разрабатывает высоконагруженный продукт;
- Осуществляет развертывание docker контейнеров и их оркестрацию;
- Выстраивает систему мониторинга и сбора метрик;
- Администрирует OS (Linux);
- Взаимодействует с командами разработчиков;

- Составляет описание серверов/решений/ подходов системе Confluence.

### **Ведущий аналитик**

- Разработка технических заданий, а также программ и методик испытаний;
- Разработка UML-диаграмм по продукту (activity, sequence);
- Формирование требований к пользовательскому интерфейсу и работе продукта;
- Формулирование процессов работы продукта;
- Составление технической документации по разработанному программному обеспечению;
- Содействие решению проблем при промышленной эксплуатации разработанного программного обеспечения в рамках своих компетенций;
- Участие в решении смежных задач ЦК НТИ в рамках своих компетенций;
- Обеспечивает аналитическую поддержку Проекта; участвует в подготовке презентаций, докладов и других материалов в рамках реализации Проекта;
- Участвует в подготовке и редактировании аналитической отчетности по Проекту.

### **Специалист службы поддержки**

- Обеспечивает организацию и контроль документооборота , в том числе с использованием ИТ-систем Института;
- Обеспечивает поддержку по оформлению, архивированию, передаче установленных форм и отчетов, информационных писем;
- Проводит опрос о проблемах в работе продукта;
- В случае необходимости подключает одного из специалистов в составе команды разработки.

## **3.4 Расположение команды разработчиков**

Проект разрабатывается на базе ЦК НТИ БСИВ Сколковского института науки и технологий.

Адрес: 121205 Москва, Территория инновационного центра ”Сколково Большой бульвар, д. 30, стр. 1, ЦК НТИ БСИВ.



Таблица 1. Этапы разработки

Этап разработки	Срок выполнения этапа
Этап «Пре-Альфа». Результаты выполнения этапа: 1. разработаны: <ul style="list-style-type: none"> <li>● автоматизированная среда построения документации;</li> <li>● автоматизированная среда сборки продукта;</li> <li>● среда тестирования продукта (далее – стенд);</li> </ul> 2. запущена работа стенда в режиме сбора и хранения данных; 3. разработаны первичные образы сервисов платформы; 4. разработана версия 1.0 документации по платформе, сервисам, развертыванию и сценариям использования.	31.12.2019
Этап «Альфа». Результаты выполнения этапа: 1. разработан слой моделей предметной области; 2. разработан модуль мониторинга работы моделей предметной области (реализация на стенде); 3. разработаны новые версии документации и новые образы платформы.	30.06.2020
Этап «Бета». Результаты выполнения этапа: 1. разработана финальная версия платформы и документации (бета-версия); 2. проведено тестирование платформы на данных внутреннего заказчика; 3. запущен(ы) пилот(ы) с индустриальным(и) партнером(-ами).	31.12.2020
Этап «Бета». Результаты выполнения этапа: 1. разработана финальная версия платформы и документации; 2. апробация и запуск в эксплуатацию версии фреймворка совместно с индустриальным партнером	31.12.2021