

**BAUM AI PLATFORM**  
**Руководство администратора**

Москва

2021

# 1 Назначение и условия применения программы

## 1.1 Виды деятельности, функции, которые автоматизированы

Программа VAUM AI PLATFORM предназначена для обработки структурированных и неструктурированных массивов данных любого типа и обучения моделей искусственного интеллекта для задач создания баз знаний, предиктивной аналитике в промышленности и медицине. Платформа позволяет выполнять полный цикл data science.

Контейнер-ориентированная архитектура платформы позволяет поддерживать модульность, функциональные блоки упакованы в соответствующие docker контейнеры. Основной модуль – конструктор искусственного интеллекта, который позволяет создавать модели искусственного интеллекта или использовать предобученные модели без необходимости прямого кодирования, по принципу drag n drop, создавая блок-схемы в нотации BPMN 2.0.

# 2 Подготовка к работе

## 2.1 Установка программы и ее обязательная начальная настройка

Для установки программы необходимо предварительно получить набор её дистрибутивов, и запустить установочный скрипт, входящий в этот набор. Также необходимо установить Python версии 3.6 и выше.

После запуска скрипта ввести первоначальные параметры программы:

- *Количество пользователей, в соответствии с лицензией на программу;*
- *Лицензионный ключ;*
- *Вариант установки – распределенный (D) или локальный на одном мощном компьютере(L);*
- *Путь установки – предоставленный ip-адрес платформы, и директорию для установки программы.*

После завершения установки на рабочем столе появится ярлык для запуска программы.

## 2.2 Общий принцип организации интерфейса

Запустить ярлык программы на рабочем столе.

Если вход выполнен успешно, то на экране появится страница интерфейса пользователя (см. Рисунок 1).

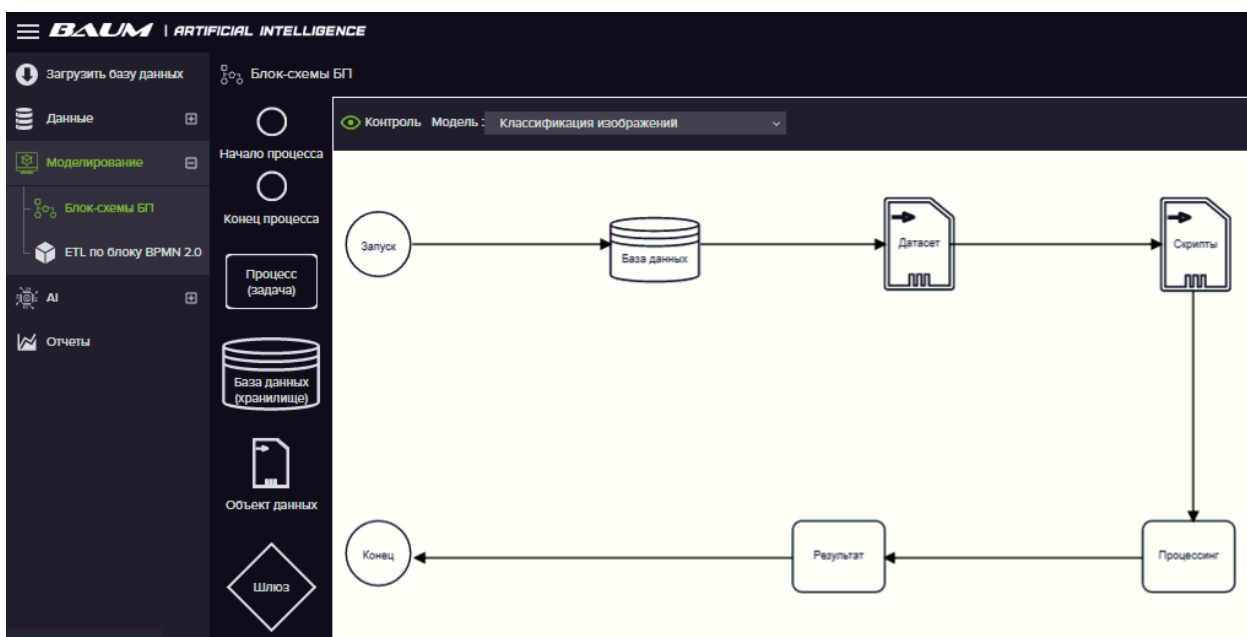


Рисунок 1 – Страница интерфейса пользователя

Таблица 1 – Составные части интерфейса

<b>Пункты меню</b>	Представлены в виде дерева вкладок. Можно легко переключаться между ними
<b>Панель объектов</b>	Представляет собой библиотеку графических элементов BPMN, из которых строится бизнес-модель. Перемещение объектов в рабочую область осуществляется по технологии «drag n drop»
<b>Панель инструментов</b>	На панели инструментов находятся кнопки для выполнения шагов бизнес-процессов, фильтры для отображения информации
<b>Рабочая область</b>	Предназначена для построения графической бизнес-модели
<b>Панель свойств</b>	Отображаются основные свойства модели, см. Рисунок 2

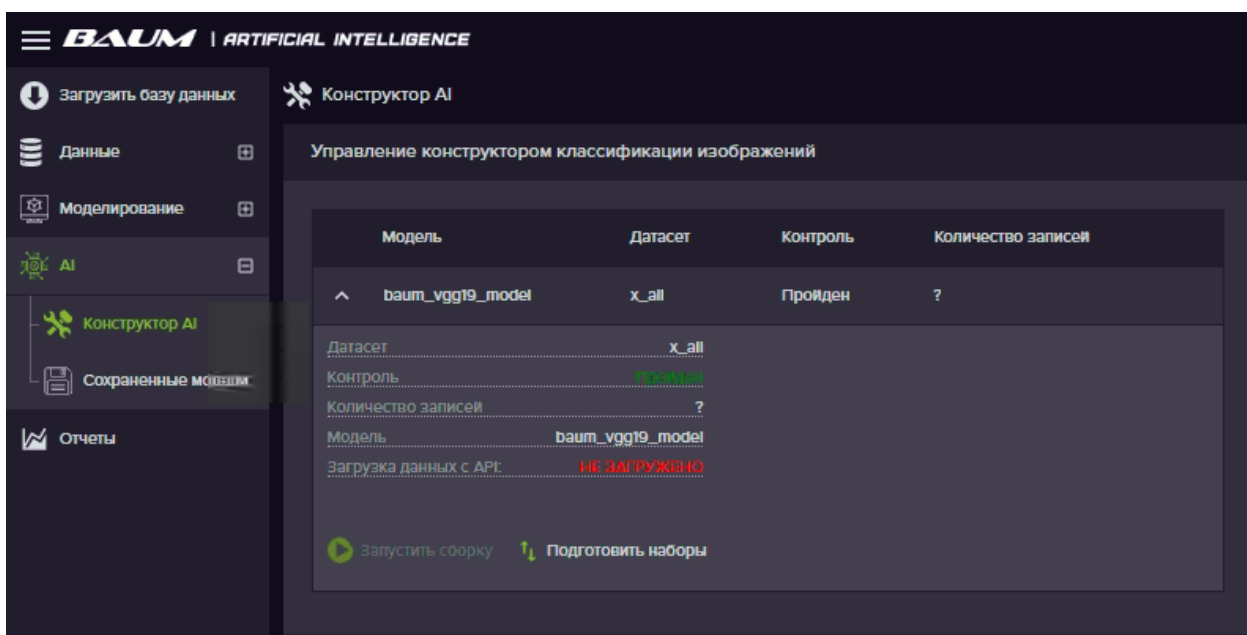






















Рисунок 2 – Панель свойств модели

Пункты меню:

	<b>Загрузить базу данных</b>	<p>Отображаются панели для загрузки входных данных, отдельно по каждому типу данных:</p> <ul style="list-style-type: none"> <li>– Тексты;</li> <li>– Изображения;</li> <li>– Временные ряды;</li> <li>– Таблицы;</li> <li>– Аудиофайлы;</li> <li>– Видеофайлы.</li> </ul>						
	<b>Данные</b>	<p>В данном пункте меню также есть возможность загрузить входные данные, отображаются загруженные датасеты (наборы данных) и все файлы, входящие в наборы.</p> <table border="1" data-bbox="644 1541 1461 1977"> <tr> <td data-bbox="644 1541 740 1832">  </td><td data-bbox="740 1541 954 1832"> <b>Тексты</b> </td><td data-bbox="954 1541 1461 1832"> <p>Текст, который нужно классифицировать (например, определить авторство), рукописный текст для распознавания, новостной фон, научные тексты «Скопус», и т.д.</p> </td></tr> <tr> <td data-bbox="644 1832 740 1977">  </td><td data-bbox="740 1832 954 1977"> <b>Таблицы</b> </td><td data-bbox="954 1832 1461 1977"> <p>Табличные данные для создания интеллектуальной базы знаний.</p> </td></tr> </table>		<b>Тексты</b>	<p>Текст, который нужно классифицировать (например, определить авторство), рукописный текст для распознавания, новостной фон, научные тексты «Скопус», и т.д.</p>		<b>Таблицы</b>	<p>Табличные данные для создания интеллектуальной базы знаний.</p>
	<b>Тексты</b>	<p>Текст, который нужно классифицировать (например, определить авторство), рукописный текст для распознавания, новостной фон, научные тексты «Скопус», и т.д.</p>						
	<b>Таблицы</b>	<p>Табличные данные для создания интеллектуальной базы знаний.</p>						

		 Временные ряды		Показатели оборудования.
		 Изображение		Фотографии анализов для постановки диагноза, изображения для классификации.
		 Аудио		Аудиофайлы для классификации, звуковая информация при наблюдении за объектами.
		 Видео		Видеопотоки мониторинга работы производства.
	<b>Моделирование</b>	Сердце программы, которое позволяет описывать бизнес-процессы, выполнять целевые действия.		
			Блок схемы бизнес-процесса	Для описания бизнес-процессов заказчика используются элементные блок-схемы, которые позволяют выстраивать цепочки, взаимосвязи, условия и т.д. Данный блок отвечает на вопрос «Что делает процесс»
			ETL по блоку BPMN 2.0	Каждый элемент в блок схеме представляет из себя карточку, в которой прописываются условия обработки информации. Откуда берутся данные, какие операции и какие преобразования выполняются над данными. Данный блок отвечает на вопрос «Как выполняется процесс». Каждая карточка содержит программный код, с содержанием основных блоков:

		<table> <tr> <td></td><td></td><td> <ul style="list-style-type: none"> <li>– Входные данные (источник входных данных);</li> <li>– Преобразование данных;</li> <li>– Выходные данные (куда и в каком виде передаются преобразованные данные)</li> </ul> </td></tr> </table>			<ul style="list-style-type: none"> <li>– Входные данные (источник входных данных);</li> <li>– Преобразование данных;</li> <li>– Выходные данные (куда и в каком виде передаются преобразованные данные)</li> </ul>			
		<ul style="list-style-type: none"> <li>– Входные данные (источник входных данных);</li> <li>– Преобразование данных;</li> <li>– Выходные данные (куда и в каком виде передаются преобразованные данные)</li> </ul>						
	AI	<table> <tr> <td></td><td>Конструктор AI</td><td>Позволяет выполнять тонкую настройку обработки данных, обучения моделей, выстраивать «pipeline» (цепочку процессов преобразования)</td></tr> <tr> <td></td><td>Сохраненные модели</td><td>Натренированные модели, которые используются в моделировании, а также модели, которые находятся в разработке</td></tr> </table>		Конструктор AI	Позволяет выполнять тонкую настройку обработки данных, обучения моделей, выстраивать «pipeline» (цепочку процессов преобразования)		Сохраненные модели	Натренированные модели, которые используются в моделировании, а также модели, которые находятся в разработке
	Конструктор AI	Позволяет выполнять тонкую настройку обработки данных, обучения моделей, выстраивать «pipeline» (цепочку процессов преобразования)						
	Сохраненные модели	Натренированные модели, которые используются в моделировании, а также модели, которые находятся в разработке						
	Отчеты	Подготовленные отчеты по установленной форме. Могут храниться результаты AI исследований, результат спроектированного процесса – документы, таблицы, графики, визуализация и т.д.						

## 3 Описание операций

### 3.1 Классификация изображений

Предварительно необходимо загрузить на локальный компьютер изображения, которые необходимо классифицировать. Под классификацией в данном сценарии понимается распознавание объектов на изображениях.

Порядок действий:

1) В главном меню перейти на вкладку **Данные** -> **Изображения**. В раздел «Загрузить данные» перенести подготовленные изображения – с помощью мыши по

технологии «drag n drop» или выбрать файлы с изображениями по одному/ набором по кнопке «Выбрать файл». Нажать кнопку «Загрузить», Рисунок 3.

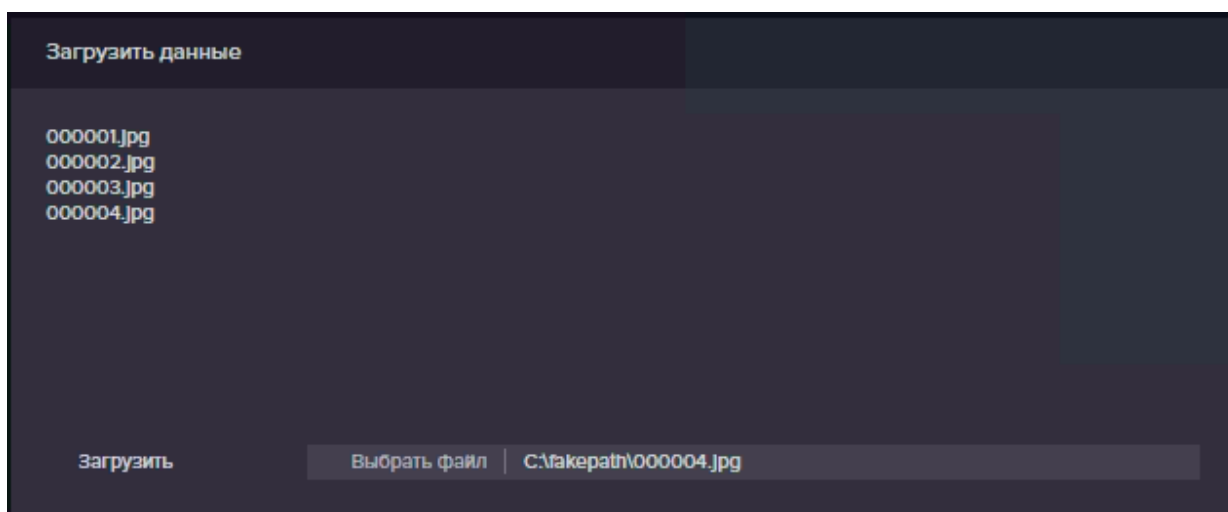


Рисунок 3 – Загрузка изображений для распознавания

Результатом выполнения данного шага является формирование датасета с выбранными изображениями, пригодного для анализа Программой, который отображается в разделе «Изображения» -> на вкладке «Наборы».

2) Перейти в пункт меню **Моделирование** -> **Блок-схемы БП** и выбрать из списка модель/ алгоритм «Классификация изображений», Рисунок 4.

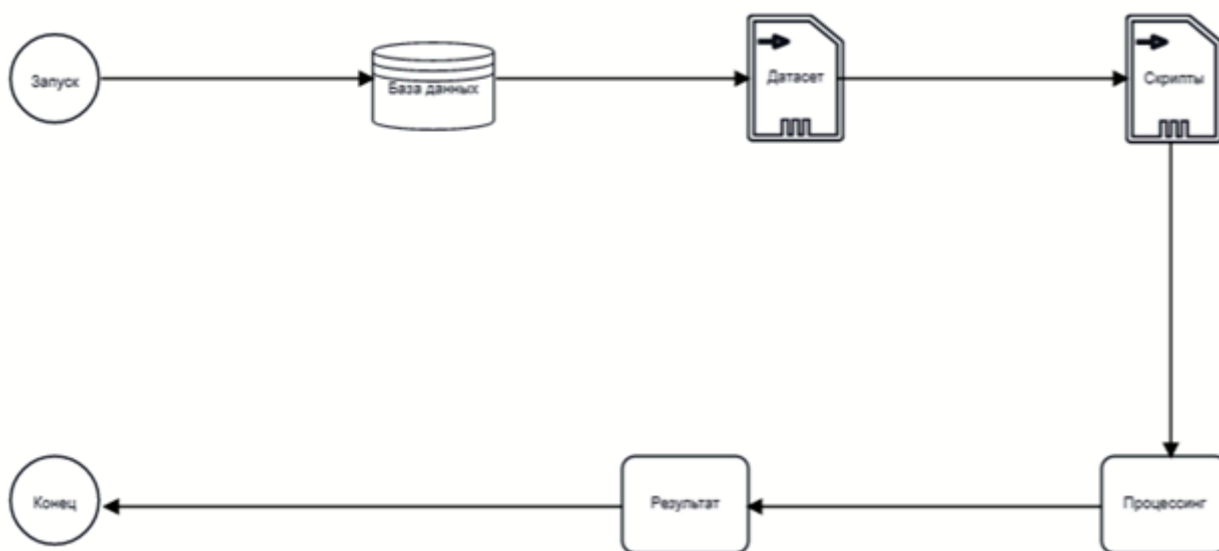



Рисунок 4 – Блок-схема классификации изображений

3) Двойным щелчком мыши открыть параметры элемента «База данных» и выбрать из списка датасет с изображениями, созданный в шаге 1. Дождаться, когда значок «Проверка совместимости параметров» отобразится в виде зеленой галочки (значок ) – это означает, что количество элементов на блок-схеме и последовательность их соединения друг с другом корректны.

4) Перейти в пункт меню **AI** → **Конструктор AI** и развернуть содержимое раздела «Управление процессом классификации изображений». Нажать кнопку «Подготовить наборы». Дождаться заполнения поля «Загрузка данных с API» значением «Загружено» – это означает, что датасет, указанный для элемента «База данных», загрузился в блок-схему, Рисунок 5.

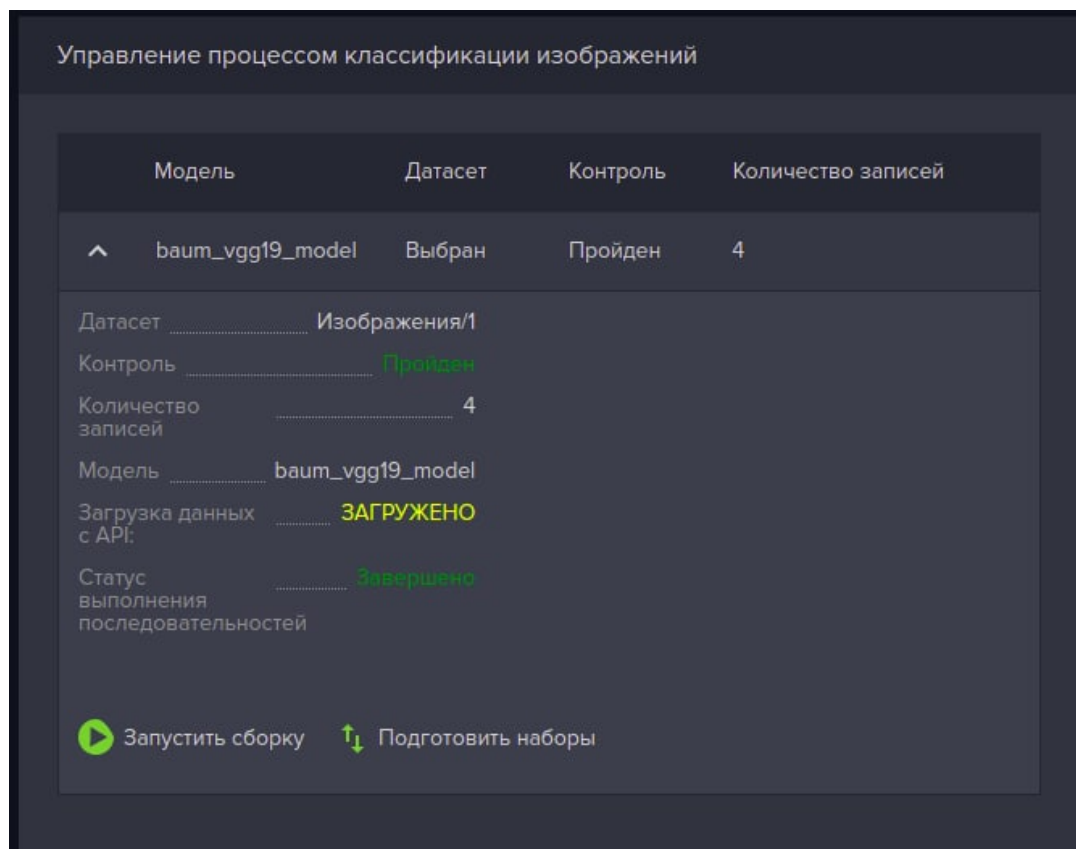


Рисунок 5 – Раздел **Управление процессом классификации изображений**

5) В разделе «Управление процессом классификации изображений» нажать кнопку «Запустить сборку». Заполнение поля «Статус выполнения последовательностей» значением «Завершено» означает, что загруженный в блок-схему датасет соответствует выбранной модели (блок-схема собрана корректно).

6) В пункте меню **Отчеты** развернуть содержимое раздела «Отчеты по классификации изображений» и проанализировать результаты распознавания загруженных изображений. Убедиться, что определения из списка с максимальной точностью соответствуют объектам на изображениях, Рисунок 6.





	
parachute 0.9999747276306152	palace 0.6660048961639404
balloon 0.000011317637472529896	castle 0.3152182996273041
crane 0.0000022459423689724645	lakeside 0.0045319413766264915
seashore 0.000001440290134269162	fountain 0.0036961298901587725
cliff 0.0000013867353345631273	mosque 0.0019089821726083755

Рисунок 6 – Результаты

### 3.2 Обучение нейронной сети для классификации текстов

Предварительно подготовить четное количество файлов в текстовом формате: 4 обучающих файла (с обозначением «train») и 4 проверочных файла (с обозначением «test»), содержащие произведения по 4 разным авторам.

Порядок действий:

1) В пункте меню **Данные** -> **Тексты** в раздел «Загрузить данные» перенести подготовленные файлы и нажать кнопку «Загрузить».

Результатом выполнения данного шага является формирование датасета с выбранными файлами.

2) Открыть пункт меню **Моделирование** – > **Блок-схемы БП**. На панели инструментов выбрать из списка модель «Обучение нейронной сети LSTM (тексты)», Рисунок 7.

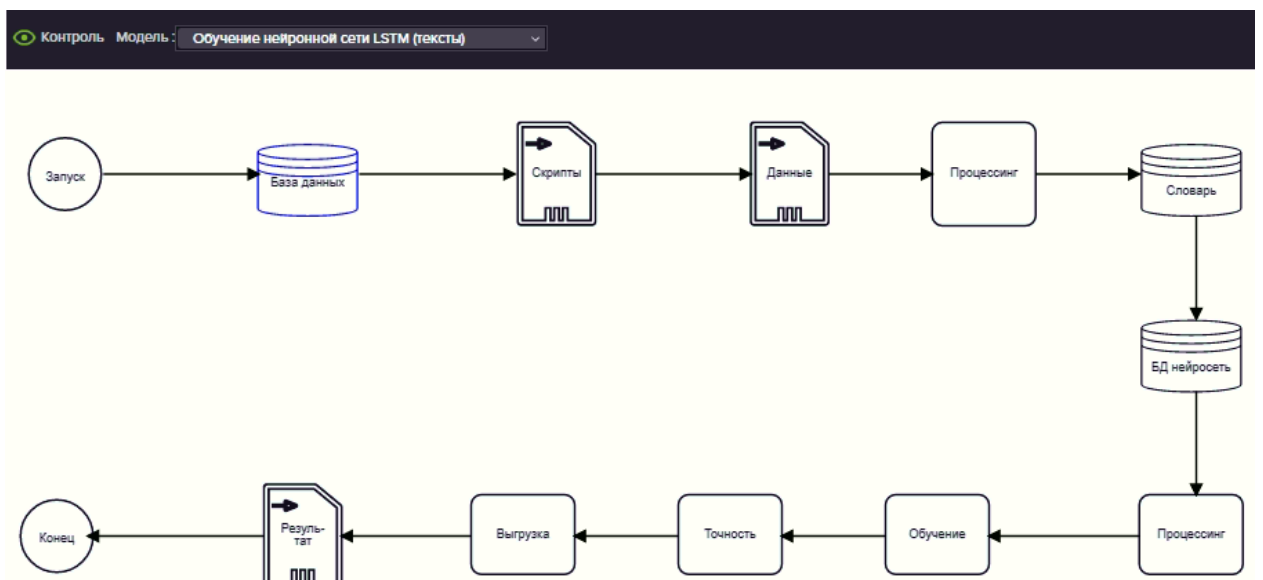


Рисунок 7 – Блок-схема обучения нейронной сети

3) Двойным щелчком мыши открыть параметры элемента «База данных» и выбрать из списка датасет с двумя наборами файлов, созданный в шаге 1. Нажать кнопку «Сохранить».



Рисунок 8 – Выбор входных параметров для сценария

Дождаться, когда значок «Проверка совместимости параметров» отобразится в виде зеленой галочки.

4) Перейти в раздел AI → **Конструктор AI**. Далее развернуть содержимое раздела «Управление процессом обучения нейронной сети LSTM (тексты)». Нажать кнопку «Начать обучение».

На экране отображается процесс обучения модели – В обучающих файлах нейронная сеть видит правильные ответы и на основании этих ответов обучается.

**Важно!** Необходимо учесть, что процесс обучения модели занимает некоторое время – до 10 минут. В это время запрещается закрывать текущий раздел Программы, иначе по завершению процесса обучения модель сохранится, но график ее обучения не отобразится в текущем разделе.

Дождаться заполнения поля «Обучение модели» значением «Обучена», Рисунок 9.

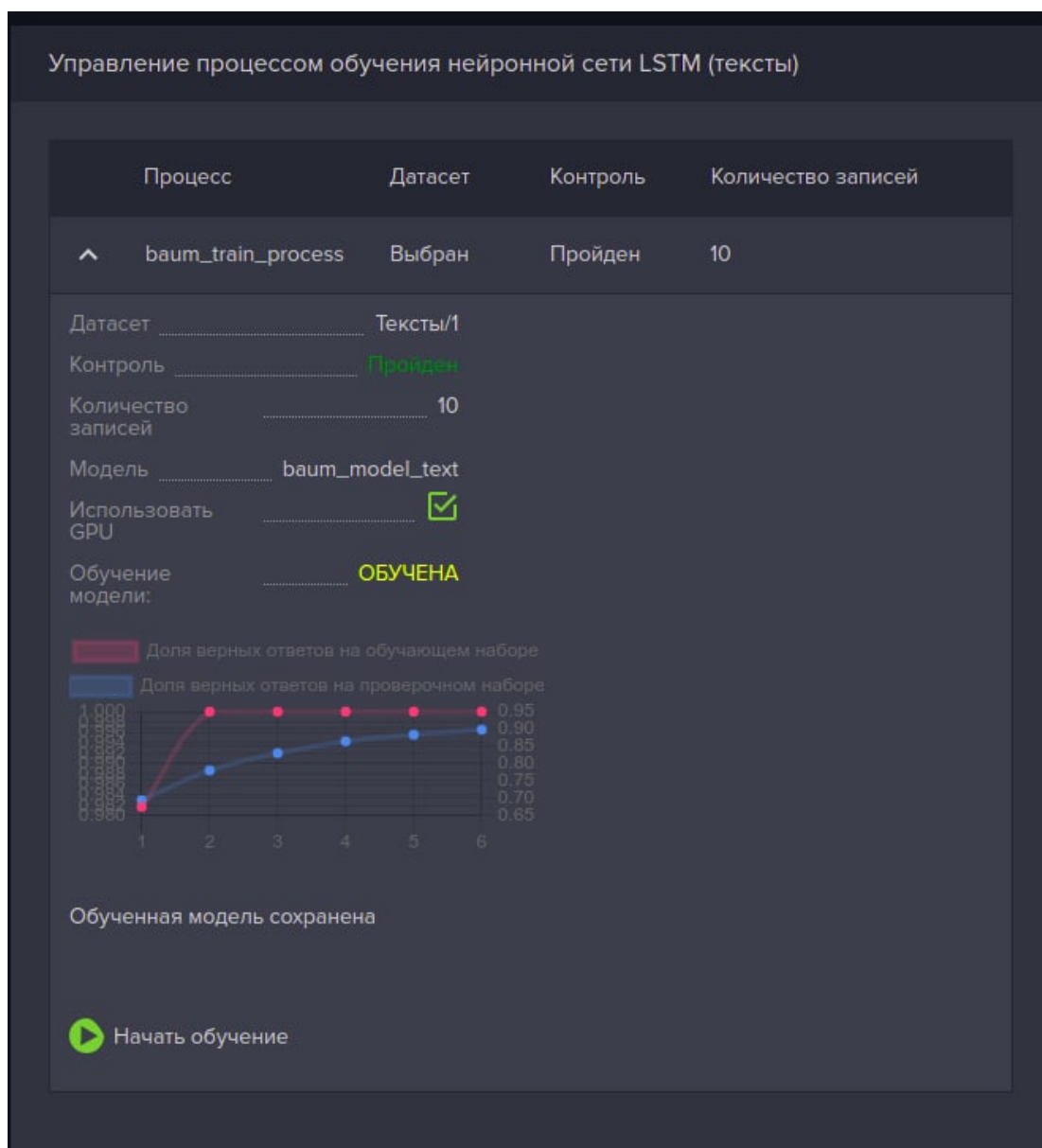


Рисунок 9 – Панель свойств, статус и визуализация обучения модели

5) Перейти в раздел AI → **Сохраненные модели**. Убедиться, что предобученная модель нейронной сети отображается в списке сохраненных моделей, Рисунок 10.

Сохраненные модели

Модель	Время сохранения	Размер файла
Bulgakov_Kilford_O-Genrl_Strugatskie_model.h5	13.04.2021, 08:46:41	49176304
Bredberri_Bulgakov_Fray_Kilford_O-Genrl_Strugatskie_model.h5	12.04.2021, 10:56:25	49181104
Bredberri_Bulgakov_Fray_model.h5	12.04.2021, 10:02:27	49173904

Рисунок 10 – Сохраненная предобученная модель

### 3.3 Кластеризация научных текстов Scopus на английском языке

«Scopus» или «Скопус» – всемирная база данных для отслеживания цитируемости статей, опубликованных в научных изданиях. База данных индексирует научные журналы, материалы конференций и серийные книжные издания, а также «профессиональные» журналы по техническим, медицинским и гуманитарным наукам.

Задача кластеризации относится к классу задач «без учителя» – Научный текст с заранее неизвестным содержимым самостоятельно разбирается на некоторое количество групп (кластеров), связанных между собой наборов ключевых слов.

Порядок действий:

- 1) Чтобы кластеризовать научный текст на английском языке предварительно нужно загрузить на Платформу файл с этим текстом в формате Excel. Сделать это можно из пунктов меню:

– **Загрузить базу данных** (Рисунок 11):

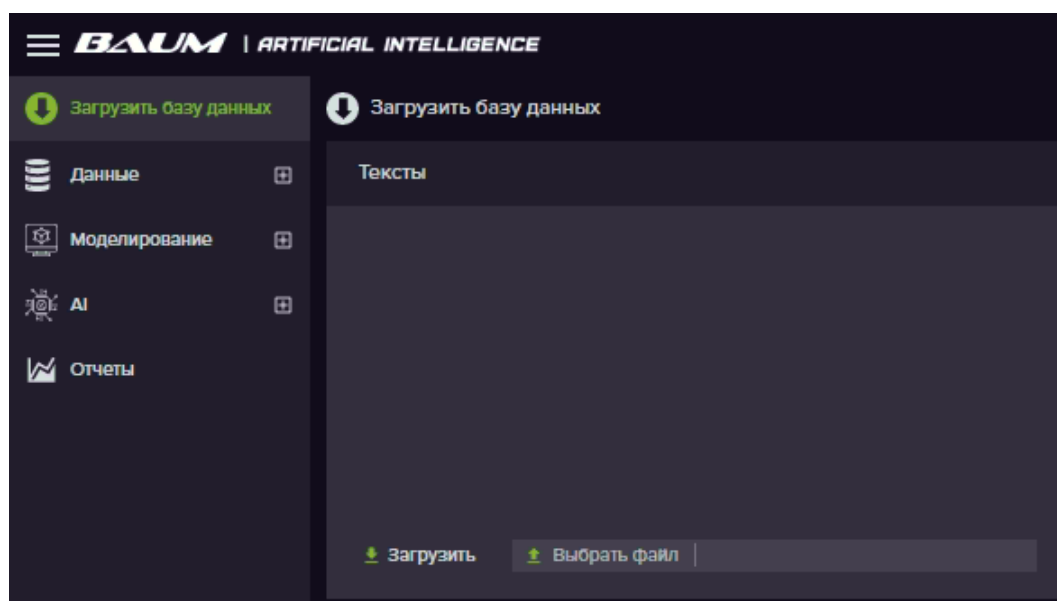


Рисунок 11 – Загрузка файла с научным текстом Scopus

– **Данные -> Тексты**. Здесь также доступен блок для загрузки файла с локального компьютера.

После загрузки файла в локальную базу Платформы он отобразится в пункте меню **Данные -> Тексты**, Рисунок 12.

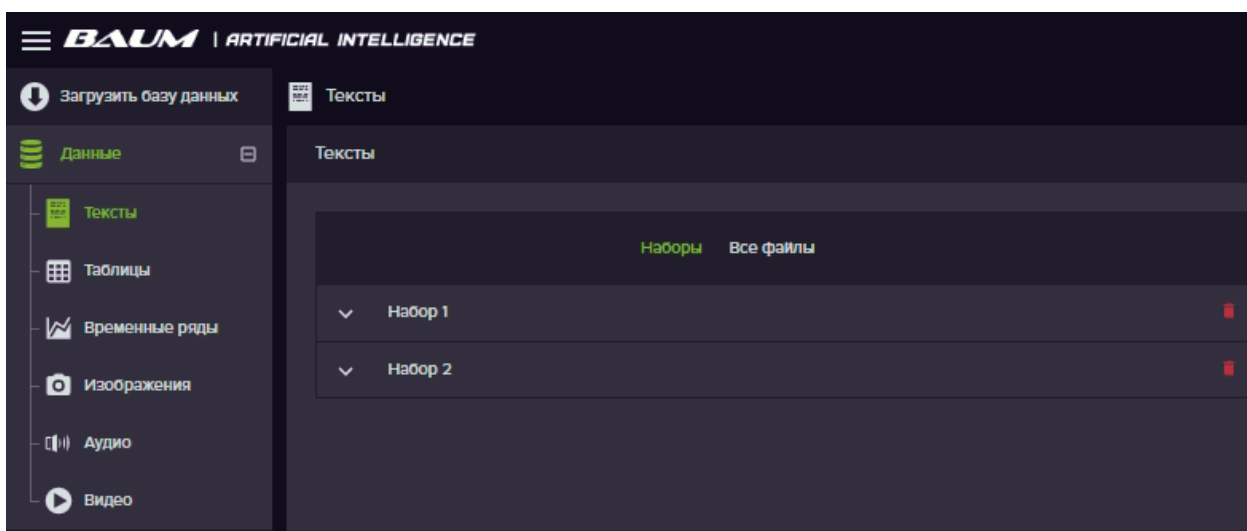


Рисунок 12 – Отображение загруженного файла в составе набора данных

- 2) Перейти в пункт меню **Моделирование** – > **Блок-схемы БП** и выбрать из списка модель «Кластеризация Scopus», Рисунок 13.

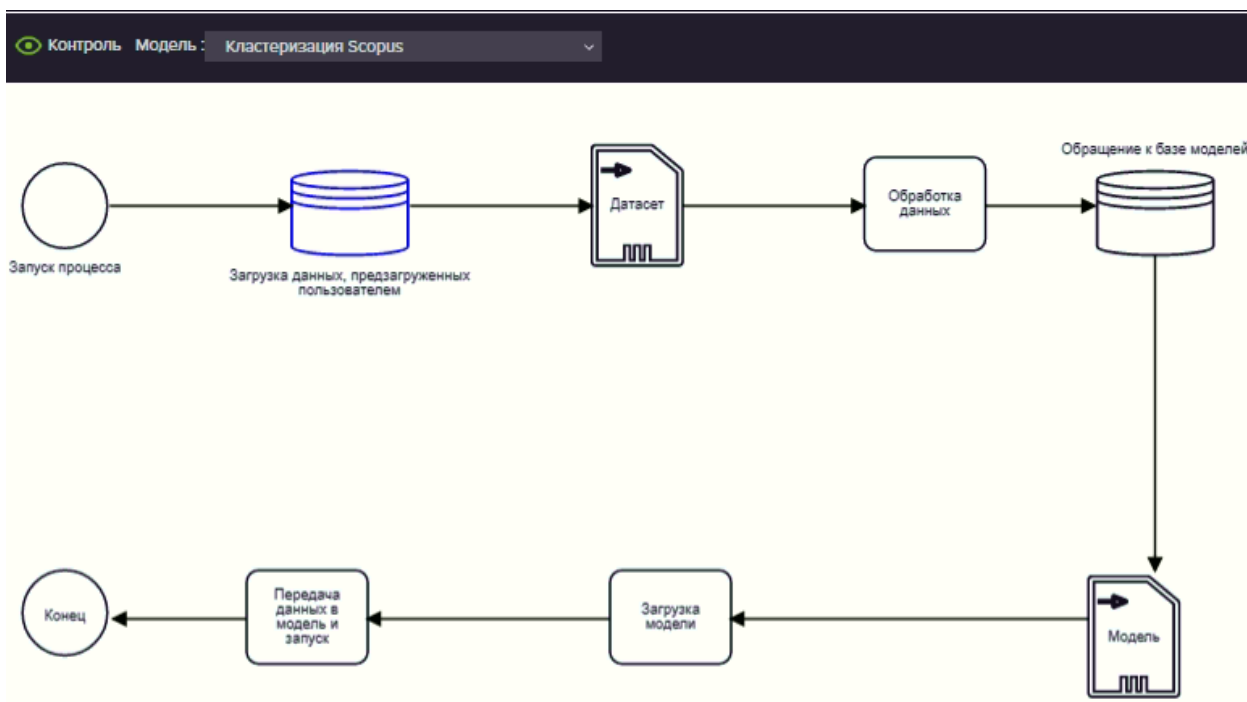


Рисунок 13 – Блок-схема кластеризации Scopus

- 3) Двойным щелчком мыши открыть параметры элемента «База данных» и выбрать из списка датасет с текстом для анализа, созданный в шаге 1. Нажать кнопку «Сохранить», Рисунок 14.

Дождаться, когда значок «Проверка совместимости параметров» отобразится в виде зеленой галочки.

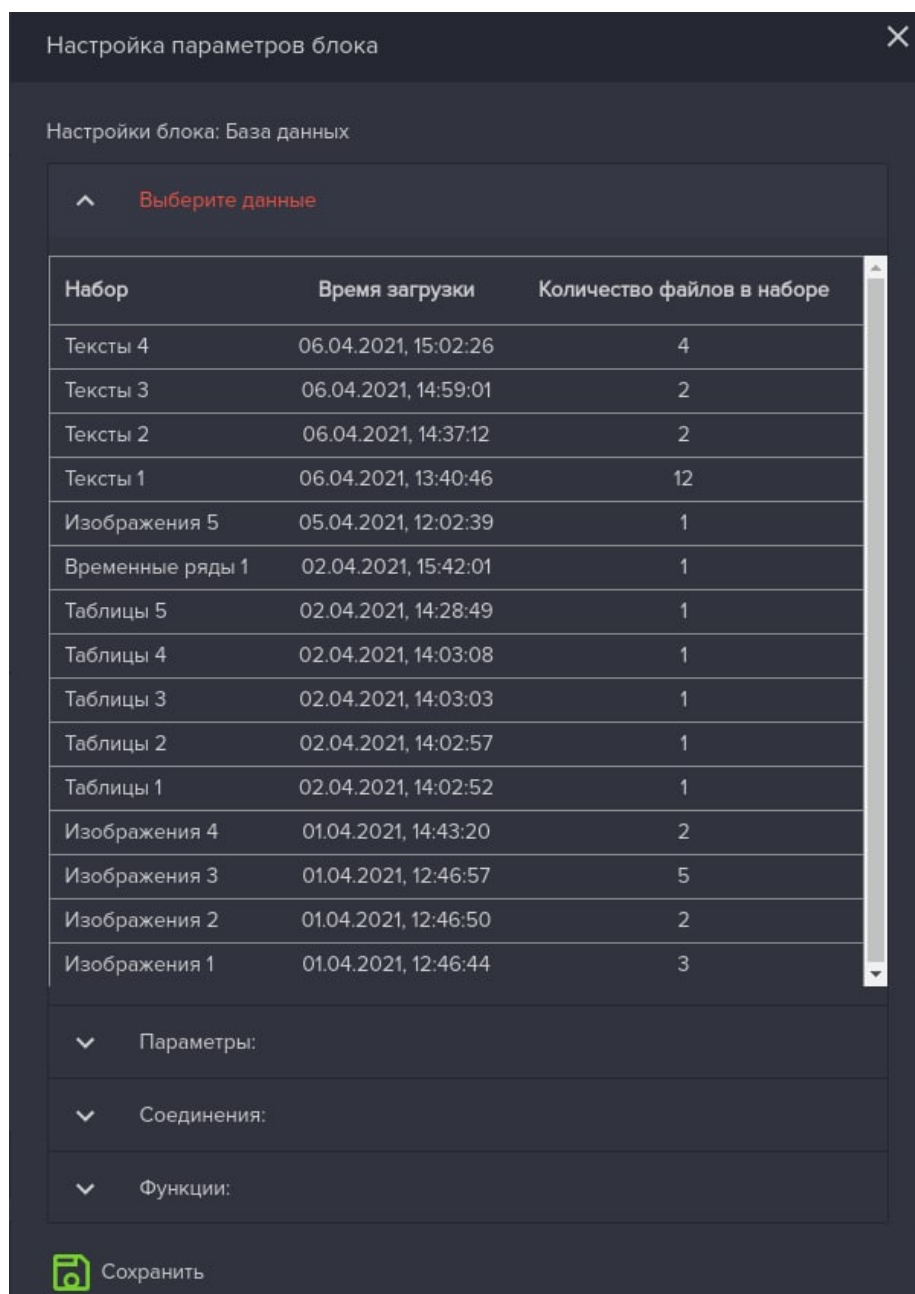


Рисунок 14 – Выбор набора данных

5) Перейти в пункт меню **AI** → **Конструктор AI** и развернуть содержимое раздела «Управление процессом кластеризации scopus данных». Нажать кнопку «Подготовить наборы», Рисунок 15.

**Важно!** Необходимо учесть, что процесс кластеризации занимает некоторое время – до 10 минут.

Дождаться заполнения поля «Загрузка данных с API» значением «Загружено».

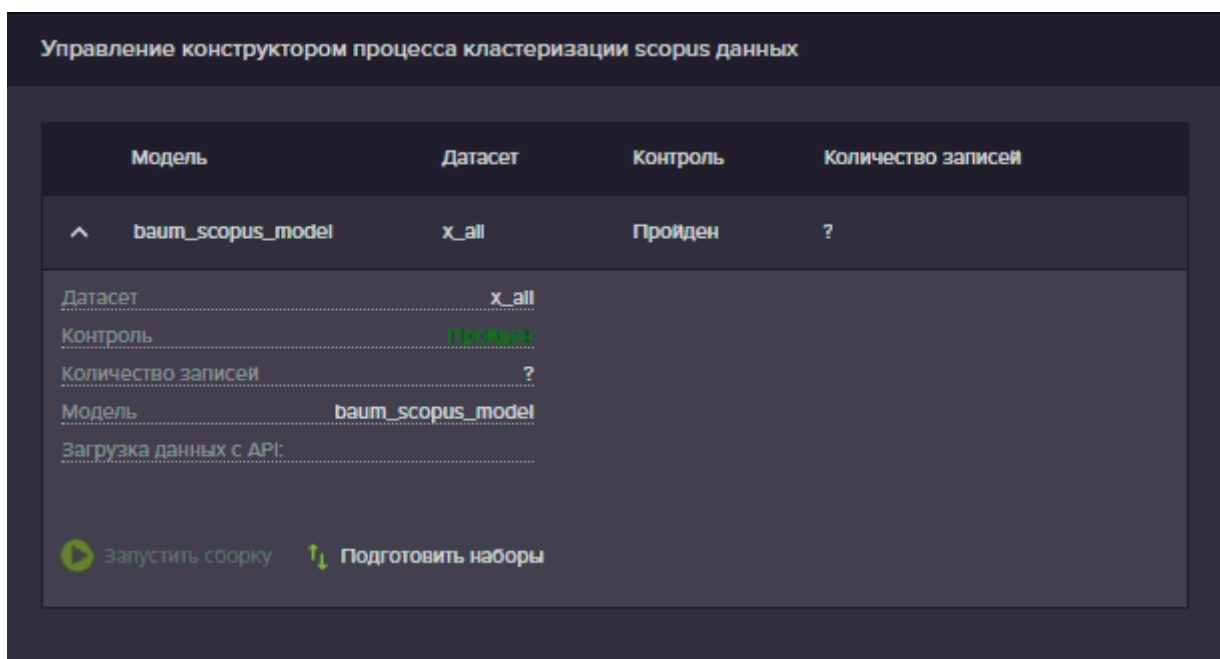


Рисунок 15 – Панель свойств модели кластеризации Scopus

6) В разделе «Управление процессом кластеризации scopus данных» нажать кнопку «Запустить сборку».

Дождаться заполнения поля «Статус выполнения последовательностей» значением «Завершено».

7) В пункте меню **Отчеты** развернуть содержимое раздела «Отчеты по модели scopus» и проанализировать результаты распознавания текста, Рисунок 16.



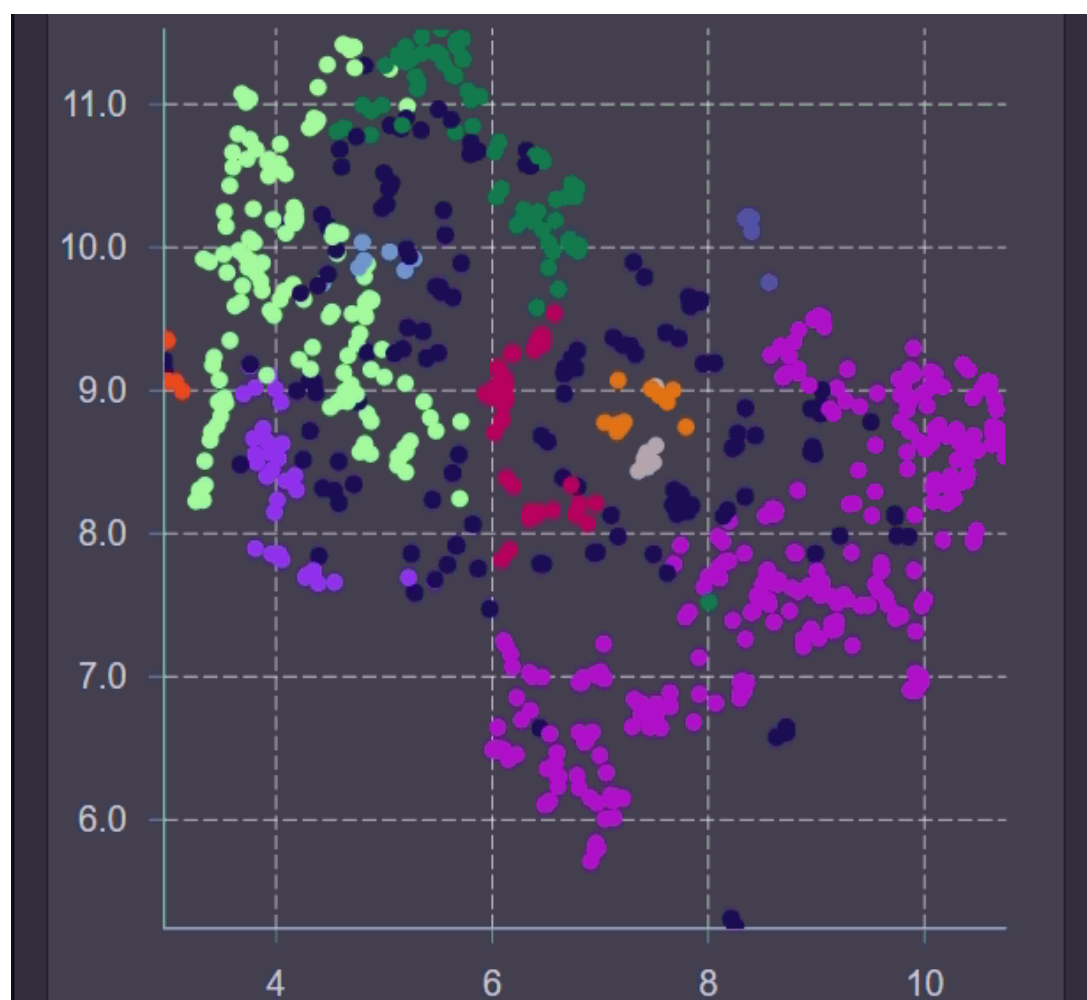
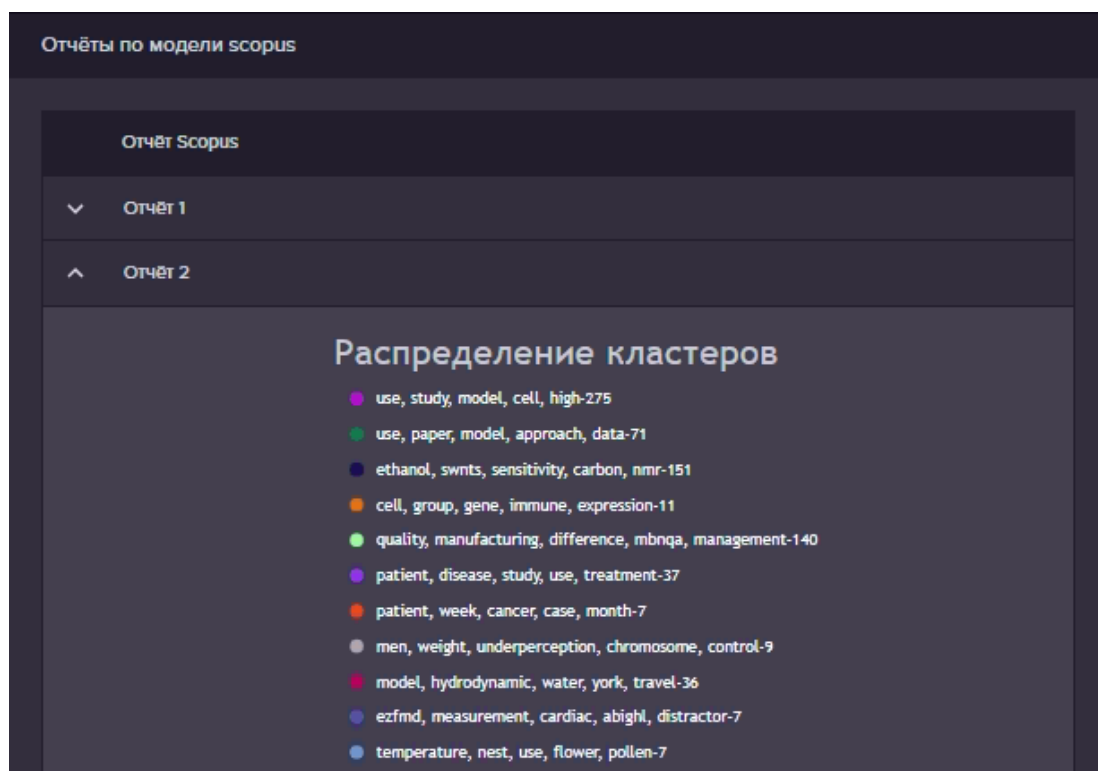


Рисунок 16 – Результаты распознавания файла с научным текстом Scopus

### 3.4 Кластеризация патентной системы на русском языке

Задачей данного сценария является анализ входящей базы с патентами с последующим выявлением и группированием патентов в кластеры по определенным признакам, которые заранее не predeterminedены и выбираются Программой автоматически.

По полученным кластерам выделяются тренды – тенденции изменения патентов (например, частоты цитируемости патентов в научных изданиях) по времени.

Предварительно необходимо подготовить базу с патентами для анализа, представленную в виде файла в формате csv – таблицу с текстовой информацией, разбитую на строки.

Порядок действий:

1) В пункте меню **Данные** -> **Таблицы** в раздел «Загрузить данные» перенести файл patents.csv, содержащий информацию с базой патентов, и нажать кнопку «Загрузить».

Результатом выполнения данного шага является формирование датасета с базой патентов.

2) В пункте меню **Моделирование** – > **Блок-схемы БП** выбрать модель «Кластеризация патентов», Рисунок 17.

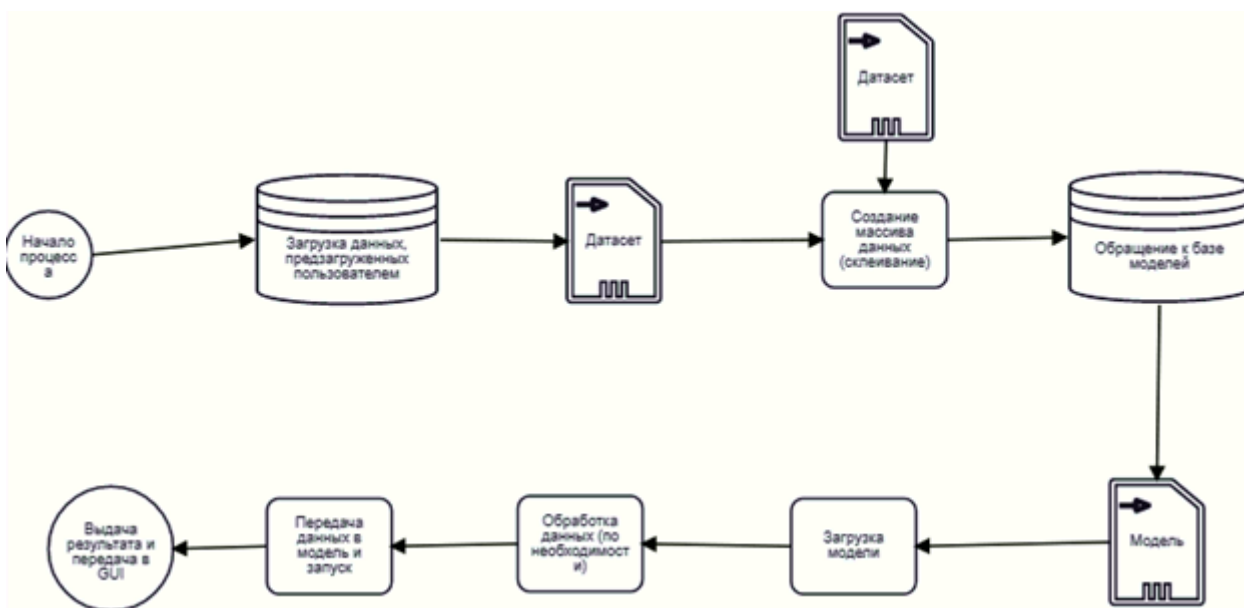


Рисунок 17 – Блок-схема кластеризации патентов

3) Двойным щелчком мыши открыть параметры элемента «Загрузка данных, предзагруженных пользователем» и выбрать из списка датасет, созданный в шаге 1. Нажать кнопку «Сохранить».

Дождаться, когда значок «Проверка совместимости параметров» отобразится в виде зеленой галочки.

4) Перейти в пункт меню **AI** → **Конструктор AI** и развернуть содержимое раздела «Управление процессом кластеризации патентов». Нажать кнопку «Обработать данные».

**Важно!** Необходимо учесть, что процесс кластеризации занимает некоторое время – до 10 минут.

Дождаться, когда поле «Получение и обработка данных» заполнится значением «Данные обработаны», Рисунок 18.

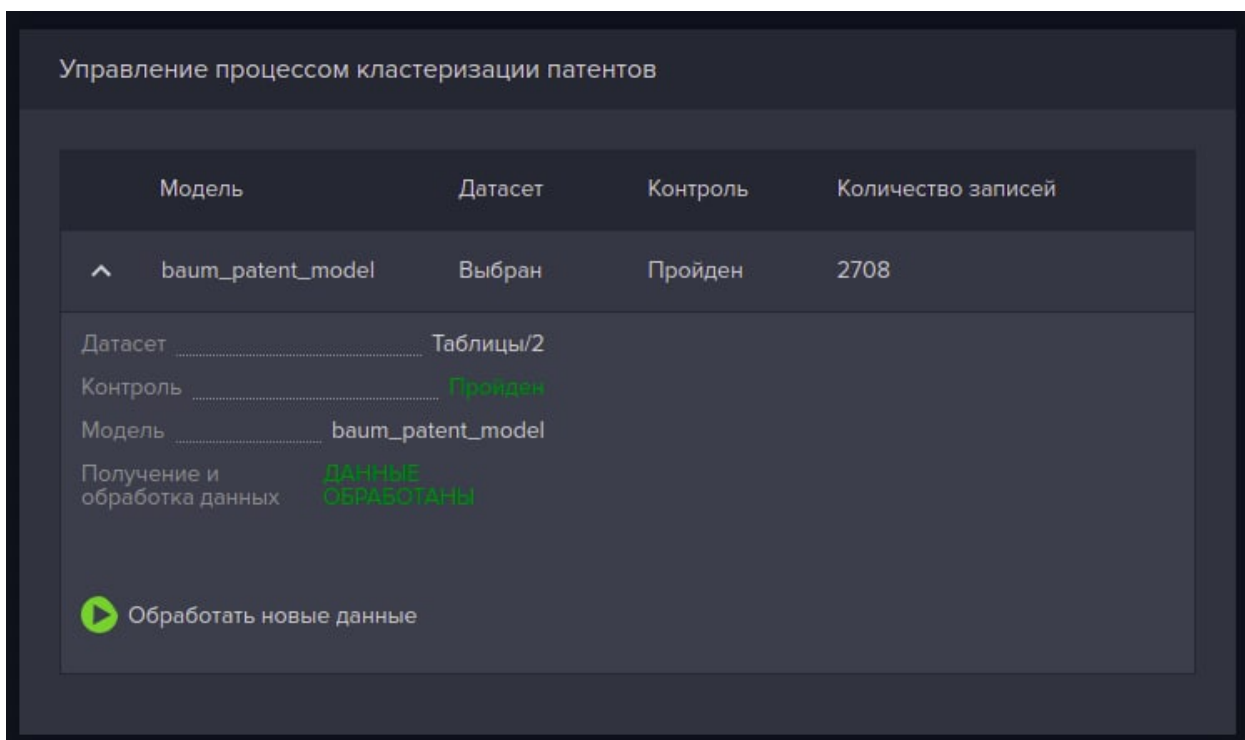


Рисунок 18 – Свойства модели кластеризации патентов

5) В пункте меню **Отчеты** развернуть отчет в разделе «Отчеты по модели кластеризации патентов» и проанализировать полученные результаты:

Сформировались кластеры с информацией о патентах, объединенные набором ключевых слов. Для каждого кластера указано количество входящих в него патентов, Рисунок 19.

Также информация о патентах представлена визуально – Кластеры перенесены на двумерную плоскость по рассчитанным уникальным координатам каждого слова во входящей базе с патентами. Используется цветовая индикация для кластеров в соответствии с рассчитанной плотностью входящих в него слов.

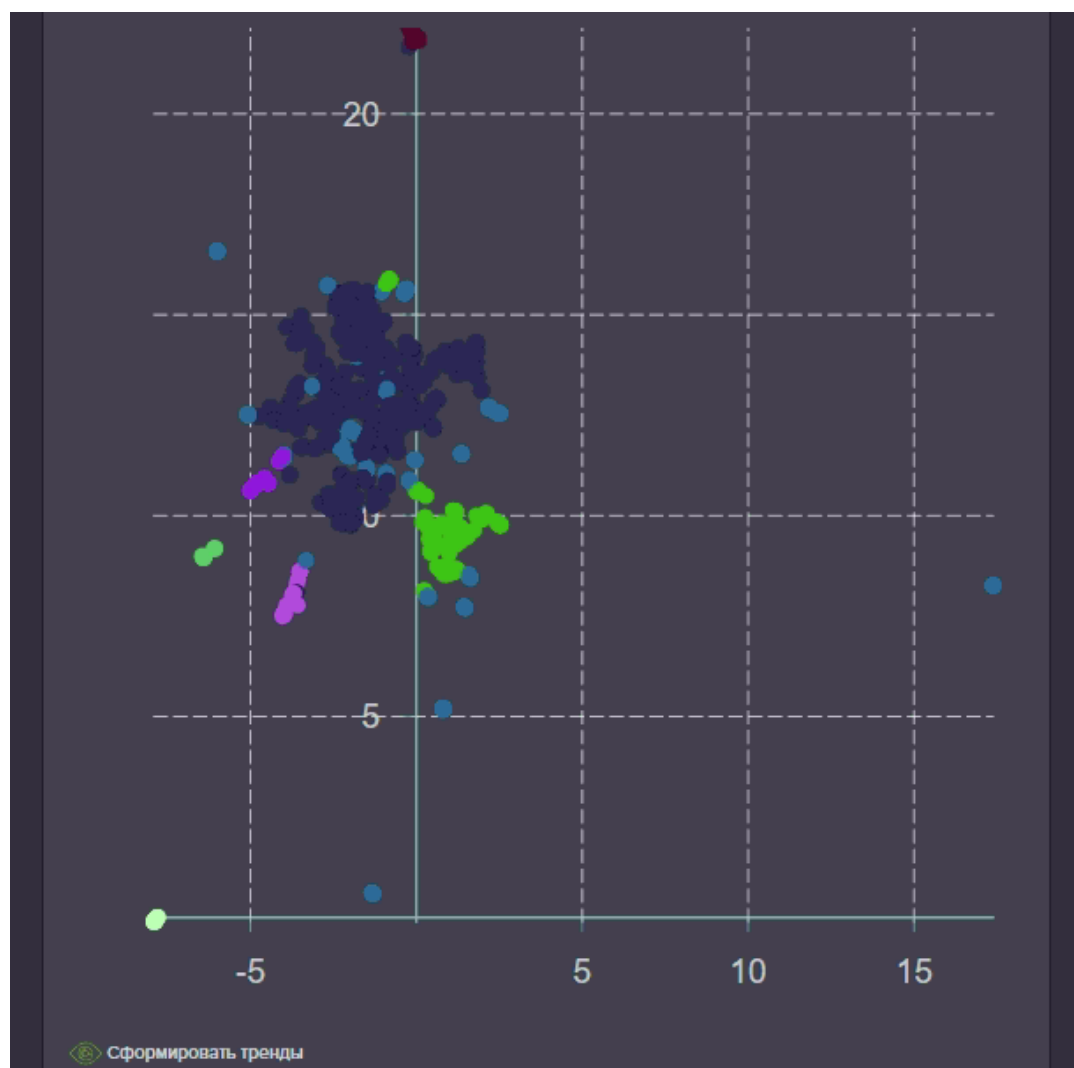
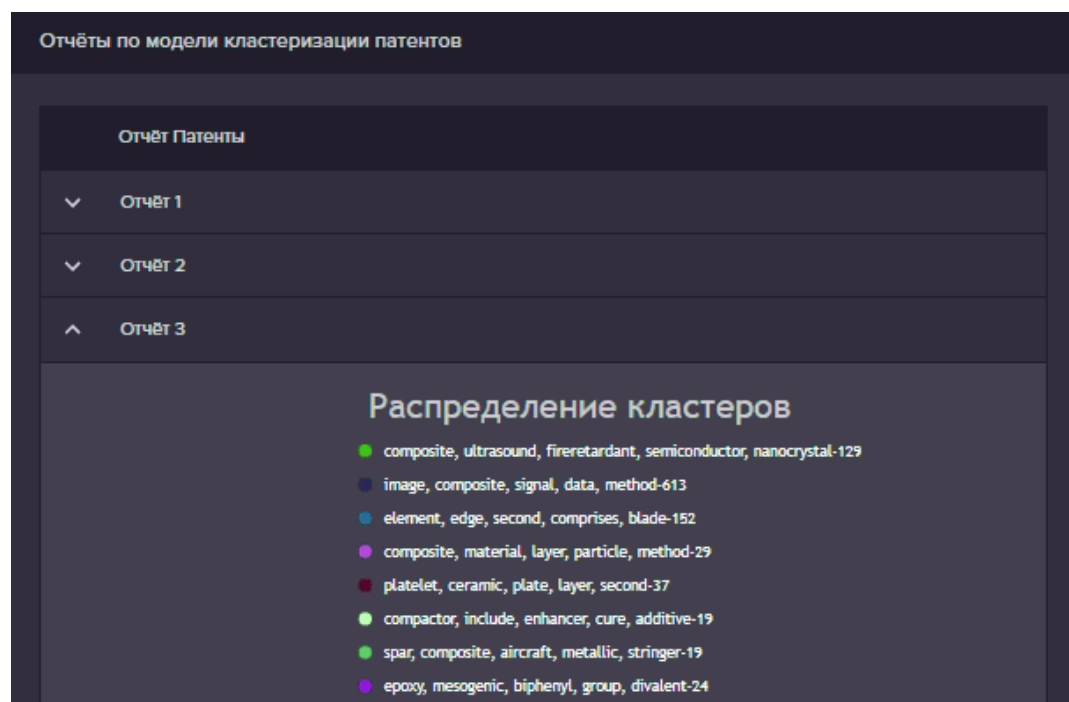


Рисунок 19 – Результаты кластеризации патентной системы

6) Рядом с графиком распределения кластеров нажать кнопку «Выделить тренды»:

По оси абсцисс отображается период времени в формате месяц, год, а по оси ординат отображается количество патентов, принадлежащих каждому из кластеров. На графике отслеживается индивидуальный итог по каждому патенту с течением времени, Рисунок 20.

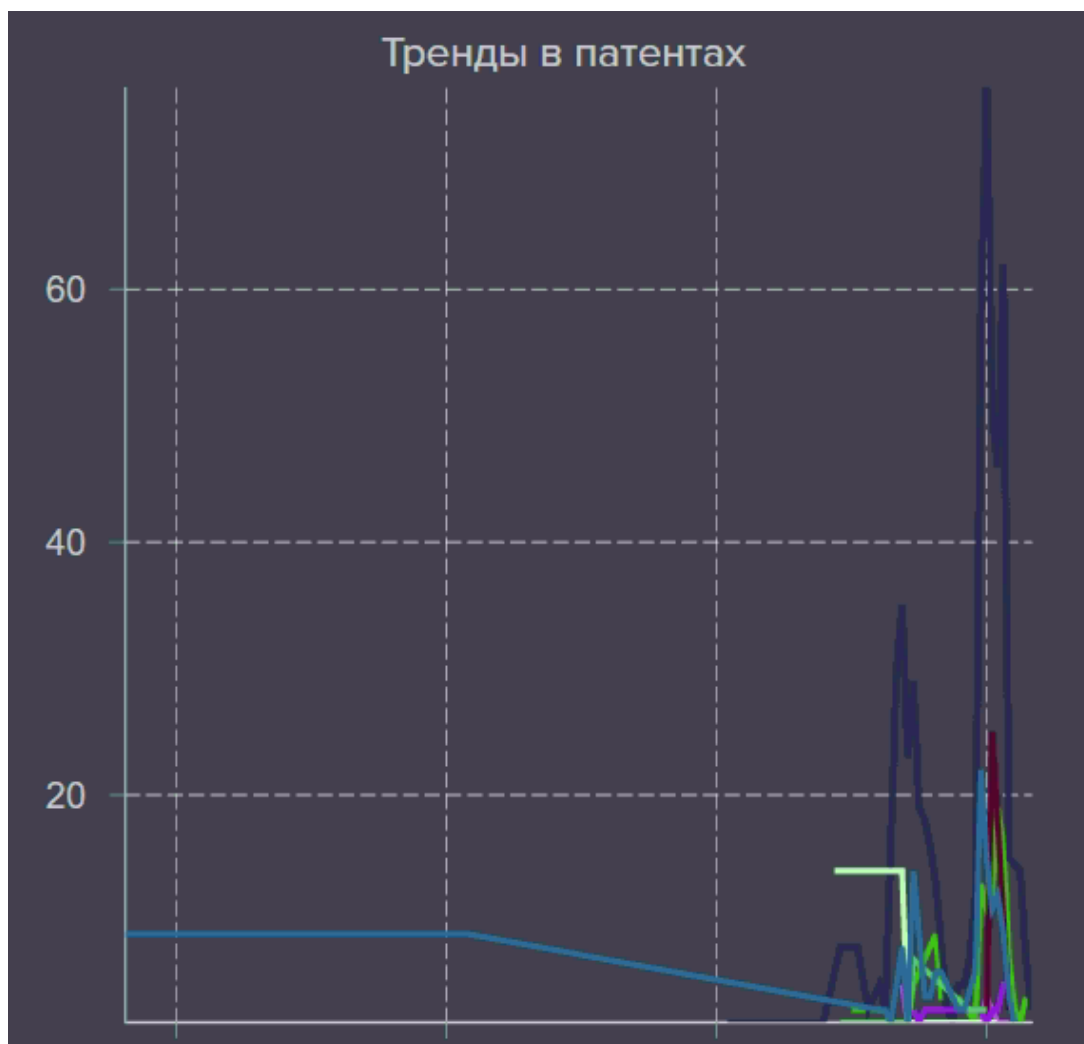


Рисунок 20 – График формирования трендов

### 3.5 Прогнозирование временного ряда

Временной ряд – это собранный в разные моменты времени статистический материал о значении каких-либо параметров (в простейшем случае одного) исследуемого процесса.

В данном сценарии программа прогнозирует изменение параметров системы на основании ранее полученных статистических данных.

Порядок действий:

1) В пункте меню **Данные** -> **Временные ряды** в раздел «Загрузить данные» перенести файл с параметрами Системы хранения данных и нажать кнопку «Загрузить».

Результатом выполнения данного шага является формирование датасета с выбранным файлом.

2) Перейти в пункт меню **Моделирование** -> **Блок-схемы БП** и выбрать из списка модель «Прогнозирование временных рядов». Откроется блок-схема, Рисунок 21.

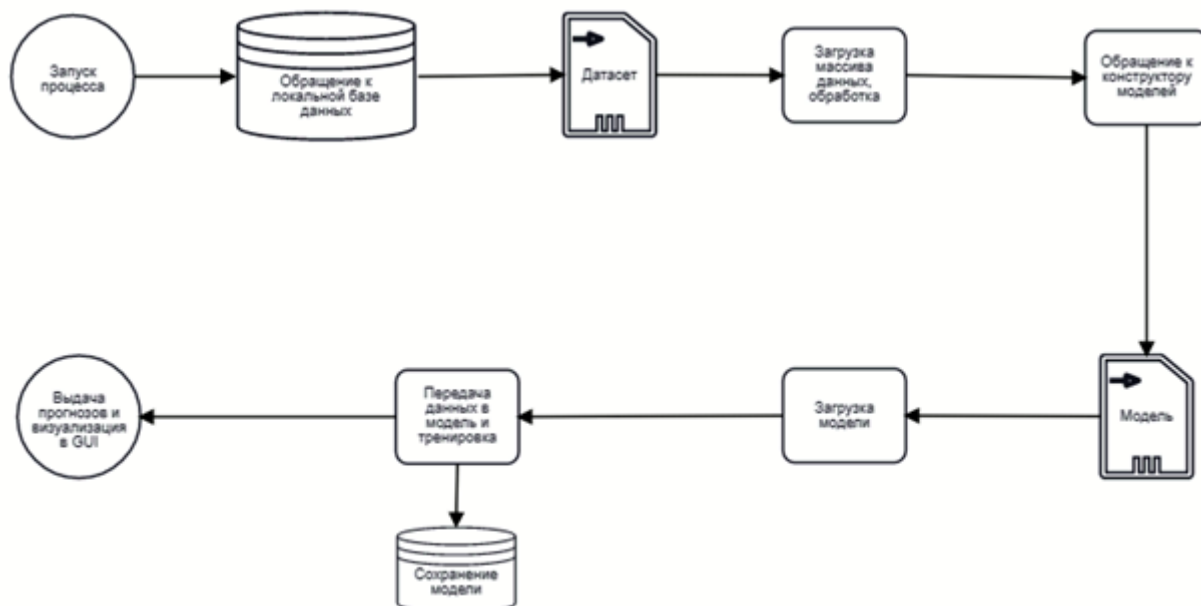


Рисунок 21 – Блок-схема прогнозирования временного ряда

3) Двойным щелчком мыши открыть параметры элемента «Обращение к локальной базе данных» и выбрать из списка датасет с файлом, созданный в шаге 1. Нажать кнопку «Сохранить».

Дождаться, когда значок «Проверка совместимости параметров» отобразится в виде зеленой галочки.

4) Перейти в пункт меню **AI** -> **Конструктор AI** и развернуть содержимое раздела «Управление прогнозированием временных рядов». Нажать кнопку «Обработать данные», Рисунок 22.

**Важно!** Необходимо учесть, что процесс кластеризации занимает некоторое время – до 10 минут.

Дождаться заполнения поля «Получение и обработка данных» значением «Данные обработаны».

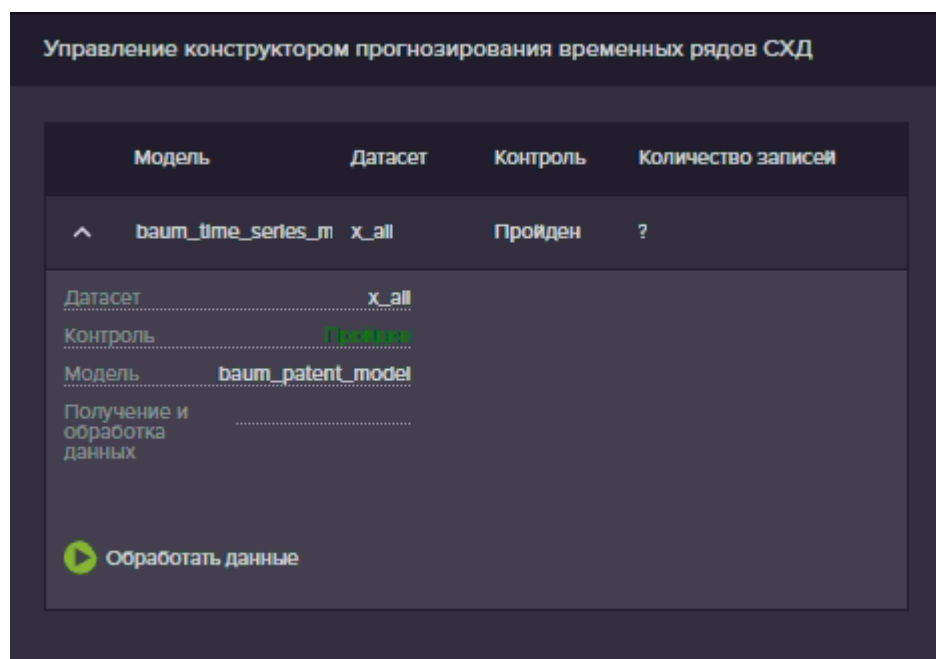


Рисунок 22 – Свойства модели

5) В пункте меню **Отчеты** развернуть содержимое раздела «Отчеты по прогнозированию временных рядов» и проанализировать результаты, Рисунок 23.

На графике отображается прогноз изменения параметров временного ряда (Предсказанные данные). Также предсказанные данные за прошедший период отображаются одновременно с реальными данными для визуализации точности прогнозирования.

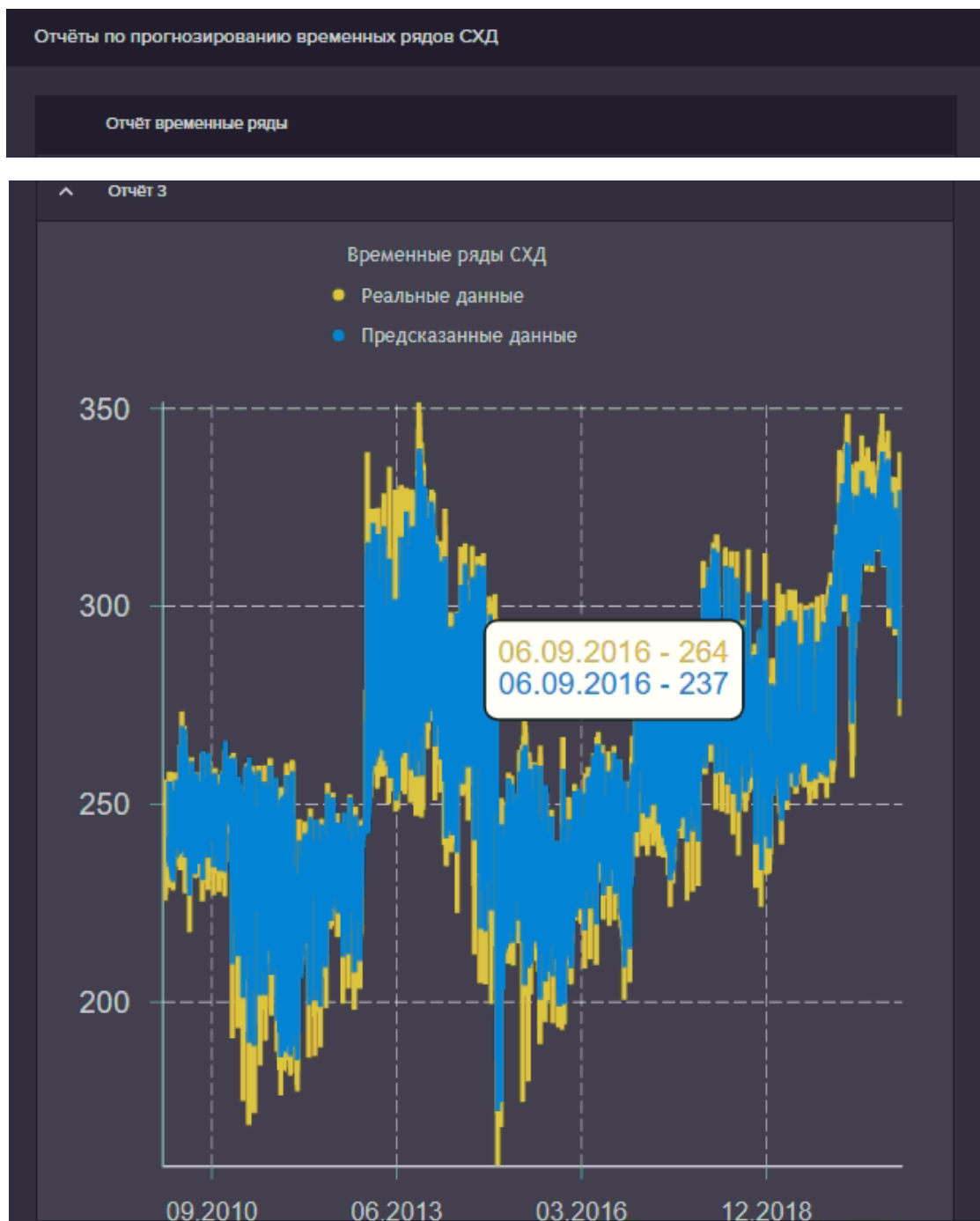


Рисунок 23 – Результат прогнозирования временного ряда

### 3.6 Прогнозирование конечных свойств композитов

В данном сценарии прогнозируются свойства композита – целевого материала, изготовленного из известных исходных компонентов, имеющих различные физические и/или химические свойства.

Порядок действий:

- 1) Перейти в пункт меню **Моделирование** -> **Блок-схемы БП** и выбрать из списка модель «Характеристика композитов», Рисунок 24.



Убедиться, что значок «Проверка совместимости параметров» отображается в виде зеленой галочки.

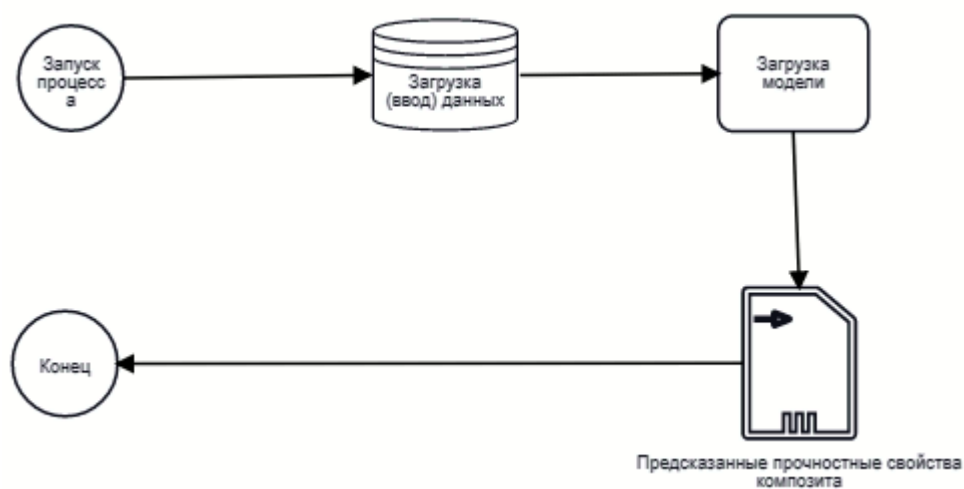


Рисунок 24 – Модель прогнозирования конечных свойств композитов

2) Перейти в пункт меню **AI** → **Конструктор AI** и развернуть содержимое раздела «Управление расчетом прочностных характеристик композитов». Сгенерировать/загрузить входные параметры для композита. Нажать кнопку «Получить данные».

Убедиться, что рассчитались и отображаются свойства композита, Рисунок 25.

Управление расчетом прочностных характеристик композитов

Модель	Датасет	Контроль	Количество записей
baum_composites	Выбран	Пройден	3

Датасет ..... Выбран

Контроль ..... **Пройден**

Выберите тип композита ..... Базальтопластик ▾

Сгенерировать данные или загрузить? ..... Загрузить ▾

Загрузка данных с API ..... **ЗАГРУЖЕНО**

Прочность на растяжение ..... 360.894 МПа5

Прочность на изгиб ..... 27.891 МПа5

Прочность на сжатие ..... 390.125 МПа5

Результат сохранения ..... **Отчет сохранен**



 Получить данные
  Сохранить результаты в отчет

Рисунок 25 – Панель свойств модели характеристик композитов

3) Нажать кнопку «Сохранить результаты в отчет». Убедиться, что сохранился отчет в пункте меню **Отчеты** -> в разделе «Отчеты по модели прочностных характеристик композитов».

Отчет содержит информацию об исходных свойствах компонентов композита и о рассчитанных прочностных характеристиках композита.

### 3.7 Распознавание текстов по предобученной модели

Подготовить файл в текстовом формате, содержащий произведение автора, для распознавания которого обучена нейронная сеть в сценарии 3.2.

Порядок действий:

1) В пункте меню **Данные** -> **Тексты** в раздел «Загрузить данные» перенести подготовленный файл и нажать кнопку «Загрузить».

Результатом выполнения данного шага является формирование датасета с выбранным файлом.

2) Перейти в пункт меню **Моделирование** -> **Блок-схемы БП** и выбрать из списка модель «Предобученная нейронная сеть LSTM (тексты)», Рисунок 26.

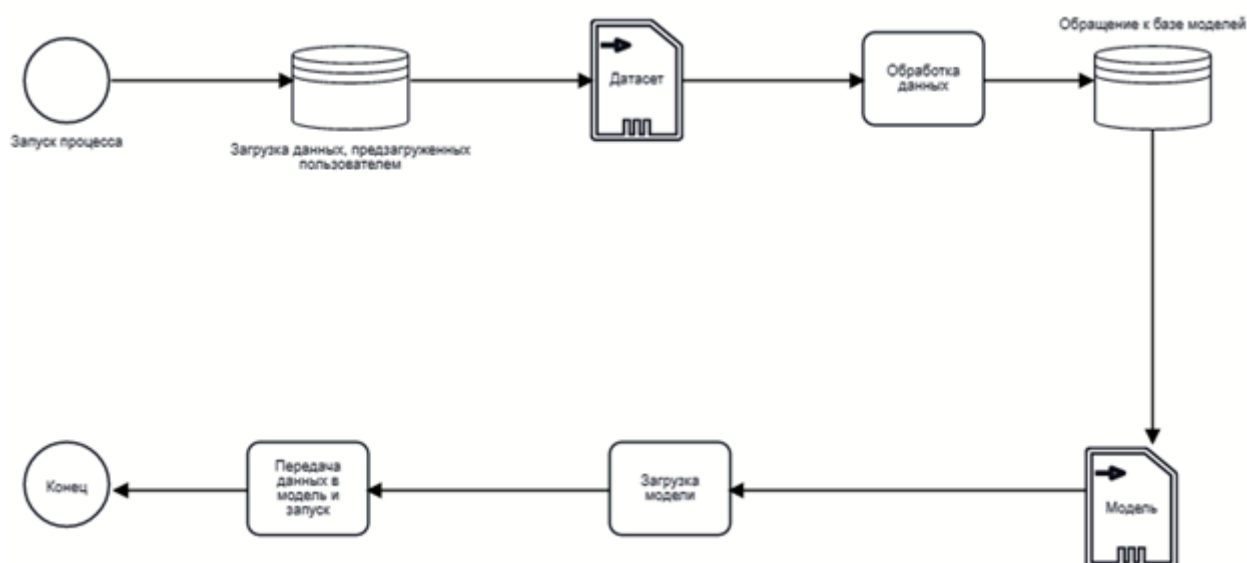


Рисунок 26 – Блок-схема распознавания текста по предобученной модели

3) Двойным щелчком мыши открыть параметры элемента «Загрузка данных» и выбрать из списка датасет с текстом для распознавания, созданный в шаге 1. Также для элемента «Обращение к базе моделей» выбрать предобученную модель нейронной сети, созданную в сценарии 3.2. Нажать кнопку «Сохранить».

Дождаться, когда значок «Проверка совместимости параметров» отобразится в виде зеленой галочки.

4) Перейти в раздел **AI** -> **Конструктор AI**. Развернуть содержимое раздела «Управление предобученной нейронной сетью LSTM (тексты)». Нажать кнопку «Получить данные».

Заполнение поля «Получение и обработка данных» значением «Данные обработаны» означает, что датасет с файлом для распознавания текста загрузился в блок-схему и соответствует выбранной модели.

В поле «Результат предсказания» отображается автор анализируемого текста, с указанием точности результата в процентах. Также результат анализа отображается графически, Рисунок 27.

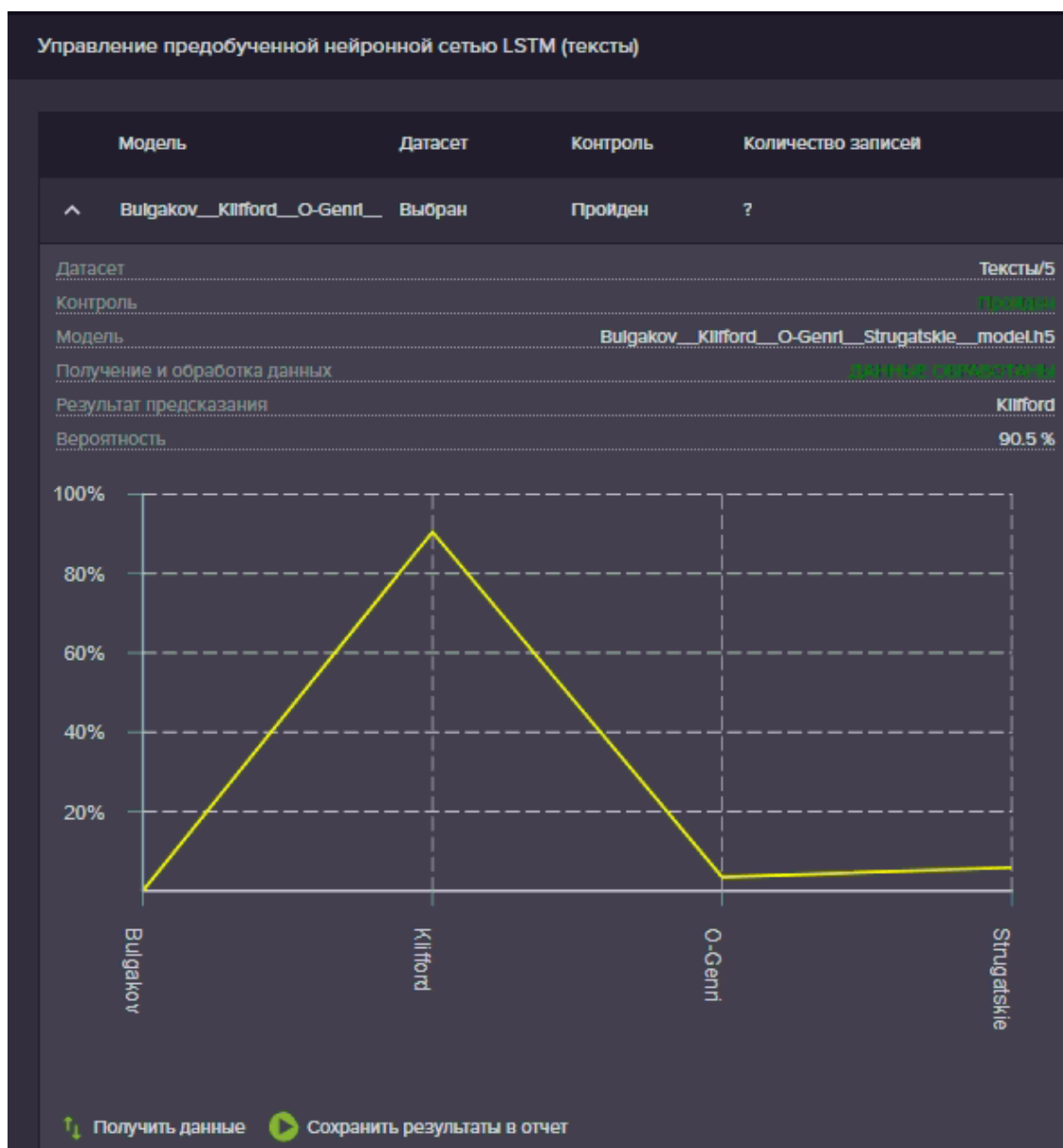


Рисунок 27 – Панель свойств модели

5) Нажать кнопку «Сохранить результаты в отчет». В пункте меню **Отчеты** развернуть содержимое раздела «Отчеты по предобученной нейронной сети LSTM (тексты)».

## ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

Сокращение	Расшифровка
AI (artificial intelligence)	Искусственный интеллект
ETL (Extract, Transform, Load)	Извлечение, преобразование, загрузка

## ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

Термин	Определение
Датасет (с англ. Data set)	Набор данных. Можно сравнить с файлом, но в отличие от файла набор данных является одновременно и каталогом, и файлом файловой системы, и не может содержать в себе другие наборы
Кластеризация	Задача группировки множества объектов на подмножества (кластеры) таким образом, чтобы объекты из одного кластера были более похожи друг на друга, чем на объекты из других кластеров по какому-либо критерию. Задача кластеризации относится к классу задач обучения без учителя
Пайплайн (с англ. Pipeline)	Последовательность стадий, внутри которых расположены задачи. Расположены они таким образом, что выход каждого элемента является входом следующего
Scopus (или Скопус)	База данных и инструмент для отслеживания цитируемости статей, опубликованных в научных изданиях