

# Похожие тексты

Настроить поиск похожих текстов

Версия 8.0



Эта документация предоставляется с ограничениями на использование и защищена законами об интеллектуальной собственности. За исключением случаев, прямо разрешенных в вашем лицензионном соглашении или разрешенных законом, вы не можете использовать, копировать, воспроизводить, переводить, транслировать, изменять, лицензировать, передавать, распространять, демонстрировать, выполнять, публиковать или отображать любую часть в любой форме или посредством любые значения. Обратный инжиниринг, дизассемблирование или декомпиляция этой документации, если это не требуется по закону для взаимодействия, запрещены.

Информация, содержащаяся в данном документе, может быть изменена без предварительного уведомления и не может гарантировать отсутствие ошибок. Если вы обнаружите какие-либо ошибки, сообщите нам о них в письменной форме.

# Содержание

<b>Настроить поиск похожих текстов</b>	<b>4</b>
1. Добавить новую модель	4
2. Настроить параметры модели	5
3. Добавить расширенные настройки	8
Результат прогноза	8

# Настроить поиск похожих текстов

ПРОДУКТЫ: [ВСЕ ПРОДУКТЫ](#)

В Creatio вы можете настраивать и обучать рекомендательные модели прогнозирования, чтобы создавать подборки похожих записей на основании анализа неструктурированных текстовых данных. К примеру, вы можете настроить автоматический подбор статей базы знаний, автоматический подбор ответов и многое другое. [Поиск похожих обращений](#) с использованием данной модели прогнозирования доступен в системе по умолчанию.

Такой метод обучения рекомендательных моделей прогнозирования называется фильтрацией на основании содержимого (content-based filtering). Он подразумевает оценку схожести по признакам содержимого определенных объектов, в данном случае — по текстовым данным. Система обобщает и ранжирует текстовые параметры субъекта и объекта прогнозирования, формируя в результате списки похожих записей, например, похожих обращений.

**На заметку.** Подробная информация о моделях машинного обучения и их использовании для решения различных бизнес-задач доступна в модульном курсе [Искусственный интеллект и машинное обучение в Creatio](#).

Формирование списка похожих объектов осуществляется в несколько этапов:

1. Настройка и обучение модели поиска похожих текстов.
2. Настройка и запуск [бизнес-процесса](#) с элементом [ *Прогнозирование данных* ].

**Важно.** Для использования функциональности предиктивного анализа данных в Creatio on-site необходимо выполнить предварительную настройку. Подробнее: [Сервис машинного обучения](#).

## 1. Добавить новую модель

**Пример.** Необходимо в разделе [ *Проблемы* ] настроить для каждой проблемы поиск похожих по теме и описанию обращений, которые пользователь в дальнейшем сможет привязать к данной проблеме.

Для этого настроим и обучим модель поиска похожих текстов.

Чтобы создать модель прогнозирования похожих текстов:

1. В рабочем месте [ *Студия* ] откройте раздел [ *Модели машинного обучения* ].
2. Нажмите кнопку [ *Добавить модель* ] —> [ *Похожесть текстов* ].
3. Заполните мини-карточку создания модели (Рис. 1):

- a. [ *Название* ] — введите название модели, по которому ее можно будет найти в реестре раздела [ *Модели машинного обучения* ] и при настройке бизнес-процесса с элементом [ *Прогнозирование данных* ].
- b. [ *Тип* ] — тип модели машинного обучения. В данном примере — “Похожесть текстов”. Поле заполняется автоматически при выборе типа модели на предыдущем шаге.
- c. [ *Искать похожее среди (Объект)* ] — выберите объект, в записях которого система будет искать похожие тексты, в данном примере — “Обращение”.
- d. [ *Искать похожее на (Субъект)* ] — укажите объект, с записями которого будут сравниваться записи, обрабатываемые при поиске похожих текстов. В нашем примере необходимо найти обращения с одинаковой корневой причиной, поэтому в данном поле необходимо указать объект “Проблема”.

Рис. 1 — Мини-карточка создания модели поиска похожих текстов

**Похожесть текстов**

Название\*  
Прогноз похожих обращений

Тип\*  
Похожесть текстов

Искать похожее среди (Объект)\*  
Обращение

Искать похожее на (Субъект)\*  
Проблема

ДАЛЕЕ ОТМЕНА

4. Сохраните мини-карточку и перейдите к настройке параметров модели поиска похожих текстов по кнопке [ *Далее* ].

## 2. Настроить параметры модели

После заполнения обязательных полей укажите параметры модели:

1. [ *Какие записи должны попасть в обучающую выборку?* ] — настройте фильтр, на основании которого система будет отбирать записи для обучения моделей. В нашем примере необходимо сузить выборку, чтобы отслеживалась похожесть текстов только в закрытых обращениях. Для этого установите следующий фильтр: “Состояние = Закрыто”.

Вы можете не указывать условия фильтрации. В этом случае для обучения будут использоваться все записи.

**На заметку.** Вы можете сохранить и обучить модель уже на этом этапе по кнопке [ *Обучить модель* ]. В этом случае результаты обучения можно будет увидеть в поле [ *Ожидаемая точность* ].

]. Чтобы сохранить результаты прогнозирования, заполните деталь [ *Настройка сохранения результатов прогноза* ].

2. [ *От каких колонок зависит прогнозируемое значение?* ] — выберите “Колонку объекта” или “Связанную колонку”, чтобы добавить колонки, по данным которых будет выполняться поиск. Например, выберите колонки объекта [ *Описание* ] и [ *Тема* ]. Для выбора доступны только текстовые колонки.
3. [ *По данным каких колонок искать похожее значение?* ] — выберите “Колонку объекта” или “Связанную колонку”, чтобы добавить колонки, для данных которых будет выполняться поиск похожих значений. Например, выберите колонки объекта [ *Решение* ], [ *Тема* ] и [ *Признаки (Симптомы)* ]. Для выбора доступны только текстовые колонки.
4. [ *Настройка сохранения результатов прогноза* ] — укажите, где в системе будет сохраняться результат прогноза. Вы можете сохранить прогноз в любом объекте системы, содержащем обязательные поля [ *Похожее для* ], [ *Похожее на* ] (тип “Справочник”) и [ *Вероятность* ] (тип “Дробное число”). Например, вы можете добавить деталь [ *Прогноз похожих обращений* ] на странице проблемы. Подробнее: [Создать новую деталь](#).

Если вы выберете данную деталь в качестве объекта прогнозирования, то названия колонок подходящего типа будут добавлены автоматически. Если таких колонок более одной, то при автозаполнении полей ниже будет указана первая из них, а остальные варианты будут доступны для выбора в выпадающем списке. Если колонок такого типа в объекте нет, то поле не заполнится. Рекомендуется проверять выбранные значения вручную перед сохранением модели.

- a. В поле [ *Объект* ] укажите объект, в котором будут храниться похожие записи. Обычно таким объектом является деталь. Обратите внимание, что указать можно уже существующий в системе объект. В нашем примере в данном поле можно указать предварительно созданную и настроенную деталь [ *Прогноз похожих обращений* ]. При выборе объекта обязательными для заполнения становятся поля [ *Похожее для* ] и [ *Похожее на* ].
- b. Поле [ *Похожее для* ] используется для определения объекта модели машинного обучения. В нашем примере в нем отображаются похожие обращения. Поле заполняется автоматически значением колонки объекта, указанного на предыдущем шаге, например “Похожее обращение”. При необходимости вы можете изменить значение, выбрав из выпадающего списка другую колонку подходящего типа. Поле является обязательным для заполнения.
- c. Поле [ *Похожее на* ] используется для определения субъекта модели машинного обучения. В нашем примере в нем отображается проблема, для которой осуществляется поиск похожих обращений. Поле заполняется автоматически значением колонки объекта, указанного на предыдущем шаге, например “Проблема”. При необходимости вы можете изменить значение, выбрав из выпадающего списка другую колонку подходящего типа. Поле является обязательным для заполнения.
- d. Поле [ *Вероятность* ] используется для ранжирования записей. В нашем примере чем больше значение, указанное в этой колонке, тем выше похожесть текста. Поле заполняется автоматически значением колонки объекта, указанного на предыдущем шаге, например “Оценка похожести”. При необходимости вы можете изменить значение, выбрав из выпадающего списка другую колонку подходящего типа. Поле является обязательным для заполнения.
- e. Поле [ *Модель машинного обучения* ] используется для указания названия модели машинного обучения, по которой осуществлялся прогноз. В нашем примере здесь будет указано название модели похожести текстов. Поле заполняется автоматически значением колонки объекта, указанного на предыдущем шаге, например “Модель машинного обучения”. При необходимости вы можете изменить значение, выбрав из выпадающего списка другую колонку подходящего типа.

Рекомендуем заполнять это поле, если вы используете несколько различных моделей прогнозирования.

- f. Поле [ *Дата прогноза* ] используется для указания даты, когда проводилось прогнозирование. Поле заполняется автоматически значением колонки объекта, указанного на предыдущем шаге, например “Дата прогноза”. При необходимости вы можете изменить значение, выбрав из выпадающего списка другую колонку подходящего типа. (Рис. 2).

Рис. 2 — Параметры модели поиска похожих текстов

The screenshot shows the 'PARAMETERS' tab of a configuration interface. It contains several sections:

- Какие записи должны попасть в обучающую выборку?** (Which records should be included in the training set?): A dropdown menu 'Действия' (Actions) is set to 'Состояние = Закрыто' (State = Closed). A checkbox 'и' (and) is checked, and there is a '+ Добавить условие' (Add condition) button.
- От каких колонок зависит прогнозируемое значение?** (Which columns does the predicted value depend on?): A list of columns is shown, including 'Описание' (Description) and 'Тема' (Topic).
- По данным каких колонок искать похожее значение?** (Which columns should be used to search for a similar value?): A list of columns is shown, including 'Решение' (Solution), 'Тема' (Topic), and 'Признаки (Симптомы)' (Features (Symptoms)).
- Настройка сохранения результатов** (Result saving settings): A table of settings for saving results.
 

Объект	Прогноз похожих обращений
Похожее для*	Похожее обращение
Похожее на*	Проблема
Вероятность	Оценка похожести
Модель машинного обучения	Модель машинного обучения
Дата прогноза	Дата прогноза

5. [ *Настройка автоматического обучения модели* ] — перетяните ползунок вправо, чтобы задать параметры автоматического переобучения модели на основании обновленных исторических данных.
- a. В поле [ *Переобучать через, дней* ] укажите длительность перерыва между обучениями модели. По истечении указанного количества дней модель будет отправлена на переобучение с использованием исторических данных, которые соответствуют настроенным фильтрам. Первое обучение модели проводится по нажатию кнопки [ *Обучить модель* ]. Если вы не хотите

переобучать модель, то оставьте поле незаполненным или введите “0”.

- b. В поле [ *Нижний порог допустимого качества* ] укажите наименьшее допустимое значение точности прогноза. При первом обучении модели это значение определит точность прогноза, по достижении которой экземпляры модели могут применяться для работы в системе. Экземпляры, не достигшие нижнего порога допустимого качества, системой не используются. Рекомендуется указывать нижний порог допустимого качества более 0,5. Точность прогноза варьируется от 0,00 до 1,00, где 1,00 — это максимально точный прогноз, а 0,00 — наименее точный. Точность прогноза рассчитывается как отношение количества правильно спрогнозированных значений к общему количеству исторических данных, на которых проводилось обучение. Подробно о механизме расчета точности прогнозов читайте в [документации Google](#).
6. В группе полей [ *Настройка фонового обновления результатов прогнозирования* ] перетащите вправо ползунок и настройте условия фильтра, если вы хотите, чтобы для выбранных записей каждый день в период минимальной загрузки системы выполнялось обновление результатов прогнозирования.

### 3. Добавить расширенные настройки

Перейдите на вкладку [ *Расширенные настройки* ], если вы хотите указать дополнительные параметры модели прогнозирования. Заполните поля аналогично тому, как это описано в настройках для модели [прогнозирования справочного поля](#), а также проверьте автоматически заполненное значение в поле, специфическом для данной модели машинного обучения: [ *Нижний порог оценки схожести* ] — самая низкая оценка схожести, при которой запись может попасть в список возможных совпадений (Рис. 5).

Рис. 3 — Расширенные параметры модели поиска похожих текстов



### Результат прогноза

В результате в Creatio будет создана новая модель, которую можно использовать для запуска бизнес-процессов, осуществляющих поиск похожих объектов в системе по неструктурированным текстовым данным.


Подробнее: [Настроить бизнес-процесс с прогнозированием](#).



В нашем примере модель поиска похожих текстов проанализирует текстовые данные объекта [ *Обращение* ], сравнит их с текстовыми данными субъекта [ *Проблема* ], после чего сформирует список похожих записей. Выборка записей для обучения будет ограничена закрытыми обращениями. Список похожих обращений будет ранжирован с учетом оценки схожести.

В результате на странице проблемы на детали [ *Прогноз похожих обращений* ] будут отображены похожие обращения (Рис. 4).

Рис. 4 — Список похожих обращений

<div>  <span>Прогноз похожих обращений</span> <span>⋮</span> </div>				
Похожее обращение <a href="#">SR00000068</a>	Оценка схожести 1,0000	Проблема Установка Microsoft Office	Дата прогноза 10.01.2021	Модель машинного обучения  <a href="#">Прогнозирование похожих обращений</a>
Описание Необходимо установить самую новую доступную версию Microsoft Office.				
Похожее обращение <a href="#">SR00000048</a>	Оценка схожести 0,3996	Проблема Установка Microsoft Office	Дата прогноза 24.01.2021	Модель машинного обучения  <a href="#">Прогнозирование похожих обращений</a>
Описание Необходимо установить пакет Microsoft Office 2013				