# Машинное обучение

## Обучение модели предсказания

Создаёт новую модель машинного обучения для предсказания значений, обучает её на основе входных данных и сохраняет итоговую модель в файл

|  |  |
| --- | --- |
| Данные для обучения | [Текст] Путь к файлу CSV, содержащему данные для обучения. Файл должен обязательно содержать корректные заголовки. Файл должен быть в формате UTF8 |
| Номер основной колонки | [Число] Номер колонки, содержащей предсказываемое значение. Нумерация начинается с нуля |
| Номера колонок данных | [Текст] Номера колонок, содержащих данные. Разделитель запятая. Нумерация начинается с нуля. Например: "1,3" |
| Номера строковых колонок | [Текст] Номера колонок, содержащих текстовые данные. Разделитель запятая. Если данное значение будет не заполнено, то тип колонки будет распознаваться автоматически. Нумерация начинается с нуля. Например: "1,3" |
| Разделитель | [Текст] Разделитель колонок CSV |
| Тип алгоритма | Выбор типа алгоритма |
| Путь к модели | [Текст] Путь сохранения файла модели |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Предсказание

Предсказывает значение на основе обученной модели машинного обучения

|  |  |
| --- | --- |
| Путь к модели | [Текст] Путь к файлу модели. Модель должна быть создана с помощью блока "Обучение модели предсказания" |
| Данные | [Таблица данных] Входные данные. Колонки у таблицы должны совпадать с колонками, на которых происходило обучение модели |
| Название целевой колонки | [Текст] Название колонки для записи результата. Если колонки нет в данных, поступающих на вход блока, то она будет добавлена. Например: "Score" |
| Результат | [Таблица данных] Результат предсказания |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Обучение модели классификации

Создаёт новую модель машинного обучения для классификации, обучает её на основе входных данных и сохраняет итоговую модель в файл

|  |  |
| --- | --- |
| Данные для обучения | [Текст] Путь к файлу CSV, содержащему данные для обучения. Файл должен обязательно содержать корректные заголовки. Файл должен быть в формате UTF8 |
| Номер основной колонки | [Число] Номер колонки, содержащей класс. Нумерация начинается с нуля |
| Номера колонок данных | [Текст] Номера колонок, содержащих данные. Разделитель запятая. Нумерация начинается с нуля. Например: "1,3" |
| Номера строковых колонок | [Текст] Номера колонок, содержащих текстовые данные. Разделитель запятая. Если данное значение будет не заполнено, то тип колонки будет распознаваться автоматически. Нумерация начинается с нуля. Например: "1,3" |
| Разделитель | [Текст] Разделитель колонок CSV |
| Тип алгоритма | Выбор типа алгоритма |
| Путь к модели | [Текст] Путь сохранения файла модели |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Классификация

Классифицирует входные данные на основе обученной модели машинного обучения

|  |  |
| --- | --- |
| Путь к модели | [Текст] Путь к файлу модели. Модель должна быть создана с помощью блока "Обучение модели классификации" |
| Данные | [Таблица данных] Входные данные. Колонки у таблицы должны совпадать с колонками, на которых происходило обучение модели |
| Название целевой колонки | [Текст] Название колонки для записи результата. Если колонки нет в данных, поступающих на вход блока, то она будет добавлена. Например: "PredictedLabel" |
| Результат | [Таблица данных] Результат классификации |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Обучение модели кластеризации

Создаёт новую модель машинного обучения для кластеризации, обучает её на основе входных данных и сохраняет итоговую модель в файл

|  |  |
| --- | --- |
| Данные для обучения | [Текст] Путь к файлу CSV, содержащему данные для обучения. Файл должен обязательно содержать корректные заголовки. Файл должен быть в формате UTF8 |
| Номера колонок данных | [Текст] Номера колонок, содержащих данные. Разделитель запятая. Нумерация начинается с нуля. Например: "1,3" |
| Номера строковых колонок | [Текст] Номера колонок, содержащих текстовые данные. Разделитель запятая. Если данное значение будет не заполнено, то тип колонки будет распознаваться автоматически. Нумерация начинается с нуля. Например: "1,3" |
| Разделитель | [Текст] Разделитель колонок CSV |
| Тип алгоритма | Выбор типа алгоритма |
| Путь к модели | [Текст] Путь сохранения файла модели |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Кластеризация

Кластеризация входных данных на основе обученной модели машинного обучения

|  |  |
| --- | --- |
| Путь к модели | [Текст] Путь к файлу модели. Модель должна быть создана с помощью блока "Обучение модели кластеризации" |
| Данные | [Таблица данных] Входные данные. Колонки у таблицы должны совпадать с колонками, на которых происходило обучение модели |
| Название целевой колонки | [Текст] Название колонки для записи результата. Если колонки нет в данных, поступающих на вход блока, то она будет добавлена. Алгоритм позволяет записать в разные колонки id кластера и расстояние. Для этого в данном свойстве укажите два названия колонок через запятую. Например: "ClusterId,Distances" |
| Результат | [Таблица данных] Результат кластеризации |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Обучение модели выявления аномалий

Создаёт новую модель машинного обучения для выявления аномалий, обучает её на основе входных данных и сохраняет итоговую модель в файл

|  |  |
| --- | --- |
| Данные для обучения | [Текст] Путь к файлу CSV, содержащему данные для обучения. Файл должен обязательно содержать корректные заголовки. Файл должен быть в формате UTF8 |
| Количество компонентов | [Число] Количество компонентов в PCA (rank). Для автоматического определения задайте значение ноль |
| Номера колонок данных | [Текст] Номера колонок, содержащих данные. Разделитель запятая. Нумерация начинается с нуля. Например: "1,3" |
| Номера строковых колонок | [Текст] Номера колонок, содержащих текстовые данные. Разделитель запятая. Если данное значение будет не заполнено, то тип колонки будет распознаваться автоматически. Нумерация начинается с нуля. Например: "1,3" |
| Разделитель | [Текст] Разделитель колонок CSV |
| Тип алгоритма | Выбор типа алгоритма |
| Путь к модели | [Текст] Путь сохранения файла модели |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Выявление аномалий

Выявление аномалий на основе обученной модели машинного обучения

|  |  |
| --- | --- |
| Путь к модели | [Текст] Путь к файлу модели. Модель должна быть создана с помощью блока "Обучение модели выявления аномалий" |
| Данные | [Таблица данных] Входные данные. Колонки у таблицы должны совпадать с колонками, на которых происходило обучение модели |
| Название целевой колонки | [Текст] Название колонки для записи результата. Если колонки нет в данных, поступающих на вход блока, то она будет добавлена. Алгоритм позволяет записать в разные колонки метку и счет. Для этого в данном свойстве укажите два названия колонок через запятую. Например: "PredictedLabel,Score" |
| Результат | [Таблица данных] Результат выявления аномалий |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Запрос к GPT модели

Этот блок позволяет отправлять запросы к классическим генеративным моделям Open AI до ChatGPT. С его помощью вы можете создавать новые тексты по запросу, выполнять разнообразные задачи классификации, суммаризации, перевода и рерайтинга текстов, написание прототипов кода на разных языках программирования, парсить полуструктурированные и неструктурированные данные, извлекать и обрабатывать факты, поддерживать диалог на разнообразные темы, и многое другое. Оплата за использование данного функционала списывается с лицевого счета клиента на платформе. Для тестовых целей каждому новому пользователю даётся возможность протестировать данный функционал без оплаты. При использовании данного функционала после окончания разработки робота необходима оплата.

|  |  |
| --- | --- |
| Запрос | [Текст] Введите запрос на естественном языке |
| Модель | Выберите модель для генерации текста. Модель "text-davinci-003" предлагает наибольшее качество генерации, однако её использование наиболее дорогое и медленное. Другие модели могут предлагать более дешевые и более быстрые результаты, за счет некоторой потери качества генерации. Методом проб и ошибок подберите модель, которая лучше всего подходит для решения ваших задач. |
| Свое название модели | [Текст] Укажите свое название модели, которую поддерживает Sherpa AI Server |
| Стоп-слова | [Текст/Список] Строка или список, содержащий не более 4 строк, которые остановят дальнейшую генерацию. Возвращаемый текст не будет содержать эти слова |
| Температура | [Число] Десятичное дробное число от 0 до 1, показывающее степень "случайности" или "креативности" результата, где 0 - наименее креативный результат, а 1 - максимально случайный. Для большинства творческих задач лучше подходит значение 0.7, а если вы хотите на один и тот же запрос каждый раз получать одинаковый ответ - установите значение 0. |
| Максимальная длина | [Число] Максимальная длина результата, выраженная в условных токенах. Для английского языка 1 токен - это 4 символа, для большинства других языков 1 токен - это 1 символ. Уменьшите это число, если хотите получать в среднем более короткие запросы, увеличьте для получения более длинных запросов. Учтите, что данное число ограничивает длину ответа, однако не обязательно полученный ответ будет указанной вами длины - в зависимости от содержимого запроса он может быть короче. |
| Автоограничение длины | При включении указанная максимальная длина длина будет автоматически скорректирована. Для этого вычисляется количество токенов в запросе и учитывается максимально возможное количество токенов для выбранной модели |
| Таймаут | [Число] Максимальное время ожидания ответа в секундах. Реальное время ожидания зависит от выбранной модели, длины вашего запроса и прогнозируемой длины ответа, а также от текущей загруженности серверов. В случае превышения установленного предела ожидания возникает ошибка. |
| Ответ | [Текст] Ответ генеративной модели на запрос |
| Итоговая длина | [Число] Итоговая длина запроса и результата (объединенных вместе), выраженная в условных токенах |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Запрос к ChatGPT

Этот блок позволяет отправлять запросы к новым генеративным моделям Open AI начиная с ChatGPT. С его помощью вы можете создавать новые тексты по запросу, выполнять разнообразные задачи классификации, суммаризации, перевода и рерайтинга текстов, написание прототипов кода на разных языках программирования, парсить полуструктурированные и неструктурированные данные, извлекать и обрабатывать факты, поддерживать диалог на разнообразные темы, и многое другое. Оплата за использование данного функционала списывается с лицевого счета клиента на платформе. Для тестовых целей каждому новому пользователю даётся возможность протестировать данный функционал без оплаты. При использовании данного функционала после окончания разработки робота необходима оплата.

|  |  |
| --- | --- |
| История чата | [Таблица Данных] Добавьте историю чата, на основе контекста которого нейросеть должна сгенерировать ответ. Указанная таблица данных должна содержать колонки с именами "role" и "content". При отсутствии таких колонок в качестве "role" будет использована первая колонка таблицы, в качестве "content" - вторая колонка. В колонке "role" можно указать только одно из следующих значений: system, assistant и user. Роль system предназначена для настройки тональности и базовых установок нейросети, например с ролью "system" можно написать "You are helpful assistant on used cars selection. You know everything on how to choose the best deal for used car". С ролью "assistant" рекомендуется включать предыдущие сообщения, сгенерированный самой нейросетью. С ролью "user" рекомендуется включать реплики, написанные пользователем-собеседником. Более старые сообщения должны быть размещены в начале таблицы, более новые - в конце таблицы. Колонка "content" должна содержать само сообщение. Данное свойство можно оставить пустым, тогда генерация будет выполнена только по содержимому свойств "Роль" и "Запрос". Если используется свойство "Инструменты", то в таблице должно быть 4 колонки: "role", "content", "tool\_call\_id", "name" |
| Запрос System | [Текст] Введите текст нового запроса к нейросети. Запрос будет выполняться с ролью system |
| Запрос User | [Текст] Введите текст нового запроса к нейросети. Запрос будет выполняться с ролью user |
| Модель | Выберите модель для генерации ответа |
| Свое название модели | [Текст] Укажите свое название модели, которую поддерживает Sherpa AI Server |
| Стоп-слова | [Текст/Список] Строка или список, содержащий не более 4 строк, которые остановят дальнейшую генерацию. Возвращаемый текст не будет содержать эти слова |
| Температура | [Число] Десятичное дробное число от 0 до 1, показывающее степень "случайности" или "креативности" результата, где 0 - наименее креативный результат, а 1 - максимально случайный. Для большинства творческих задач лучше подходит значение 0.7, а если вы хотите на один и тот же запрос каждый раз получать одинаковый ответ - установите значение 0. |
| Максимальная длина | [Число] Максимальная длина результата, выраженная в условных токенах. Для английского языка 1 токен - это 4 символа, для большинства других языков 1 токен - это 1 символ. Уменьшите это число, если хотите получать в среднем более короткие запросы, увеличьте для получения более длинных запросов. Учтите, что данное число ограничивает длину ответа, однако не обязательно полученный ответ будет указанной вами длины - в зависимости от содержимого запроса он может быть короче. |
| Автоограничение длины | При включении указанная максимальная длина длина будет автоматически скорректирована. Для этого вычисляется количество токенов в запросе и учитывается максимально возможное количество токенов для выбранной модели |
| Инструменты | [Таблица Данных] Таблица с описанием инструментов, которые может выбрать нейросеть для решения задачи вместо текстового ответа. Этот функционал поддерживают не все модели. Таблица данных должна содержать три колонки: "name" - название инструмента (латинские буквы и цифры, не более 64 символов), "description" - описание инструмента (может быть на отличных от английского языках), "parameters" - аргументы передаваемые в инструмент при его вызове в формате JSON. Пример JSON-описания параметров: "{\"type\": \"object\", \"properties\": { \"query\": { \"type\": \"string\", \"description\": \"Request for a Google search result\" } }, \"required\": [\"query\"] }". При вызове инструмента нейросеть выберет подходящий инструмент из числа доступных, опираясь на указанное вами описание инструментов, а также заполнит (сгенерирует) значения указанных вами параметров для вызова этих инструментов. Более подробно про механизм использования инструментов можно узнать из описания выходного свойства CalledTools. |
| Использовать инструмент | [Текст] При необходимости можно принудительно указать название инструмента, который нейросеть должна использовать в текущем запросе. Название должно совпадать с названием одного из инструментов из колонки "name" в таблице "Инструменты". |
| Таймаут | [Число] Максимальное время ожидания ответа в секундах. Реальное время ожидания зависит от выбранной модели, длины вашего запроса и прогнозируемой длины ответа, а также от текущей загруженности серверов. В случае превышения установленного предела ожидания возникает ошибка. |
| Роль ответа | [Текст] Роль, с которой ответила нейросеть |
| Ответ | [Текст] Текст ответа нейросети |
| Итоговая длина | [Число] Итоговая длина запроса и результата (объединенных вместе), выраженная в условных токенах |
| Вызываемые инструменты | [Список] Список объектов ToolItem. Если нейросеть после запроса вместо текстового ответа предлагает использовать один или несколько инструментов, указанные в свойстве "Инструменты", то в данном свойстве будет возвращен список объектов ToolItem. Объект содержит следующие свойства: id, name, arguments. Значения id, name необходимо использовать в таблице "История чата". Свойство arguments является словарем, где ключ - это название параметра вызываемого инструмента, а значение является сгенерированным нейросетью значением этого параметра. При использовании механизма инструментов после вызова данного блока вы должны проверить, выбрала ли нейросеть какой-либо инструмент, по содержимому данного списка. Если список не пустой, вы должны самостоятельно реализовать функционал выбранного инструмента или инструментов в сценарии робота, а затем в следующем вызове данного блока вернуть нейросети результат выполнения инструмента, который нейросеть будет использовать в генерации последующих сообщений. Результат выполнения инструментов необходимо записывать в таблицу "История чата" следующим образом: в колонку "role" записывается константное значение "tool", в колонку "content" - значение результата выполнения инструмента (который вы реализовали в вашем сценарии и теперь сообщаете этот результат нейросети), в колонку "tool\_call\_id" - значение текущего ToolItem.id, в колонку "name" - значение текущего ToolItem.name. |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Подсчитать токены

Подсчитывает количество токенов в указанном запросе

|  |  |
| --- | --- |
| Запрос | [Текст] Запрос для подсчитывания количества токенов |
| Модель | Выбор модели |
| Количество токенов | [Число] Возвращается количество токенов в запросе |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Получить эмбеддинги

Блок позволяет получить эмбеддинги по строке или списку строк с использованием сервиса Open AI. Эмбеддинг - это вектор (набор чисел), которые характеризуют смысл, ассоциированный с переданной входной текста. Слова или предложения с близким смыслом будут иметь эмбеддинги с минимальным косинусным расстоянием. Эмбеддинги в том числе могут использоваться для поиска наиболее близких по смыслу слов, строк или абзацев в базах данных документов.

|  |  |
| --- | --- |
| Текст | [Текст/Список] Входной текст или список текстов, для которого будут рассчитаны эмбеддинги. Если на вход передан список текстов, на выходе также будет список эмбеддингов для каждого из элементов входного текста. Максимально допустимая длина каждой строки - 8192 токенов. |
| Модель | Выберите модель для генерации ответа |
| Как таблица данных | При включении результат будет возвращен как таблица данных с двумя колонками: "Text" и "Embeddings" |
| Таймаут | [Число] Максимальное время ожидания ответа в секундах. |
| Эмбеддинги | [Текст/Список/Таблица Данных] Если на вход была передана одна строка, то выход - это векторное представление эмбеддинга в формате JSON (массив чисел). Если на вход был передан список строк, то на выходе будет список эмбеддингов в формате JSON (массив чисел). Если включено свойство "Как таблица данных", то на выходе будет таблица данных с двумя колонками: "Text" и "Embeddings" |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Найти эмбеддинги

Блок позволяет по заданному запросу-эмбеддингу найти в базе знаний, представляющей собой таблицу объектов и их эмбеддингов, самые похожие строки и возвратить индексы этих строк в виде списка, отсортированного по убыванию похожести. Эмбеддинги и для запроса, и для таблицы эмбеддингов, по которой осуществляется поиск, могут быть получены с помощью блока Получить эмбеддинги. При первом использовании блока происходит построение базы данных векторов и это может занимать длительное время. При последующих вызовах блока с одним и тем же "ID базы" будет использоваться уже созданная база данных векторов и скорость работы блока возрастает многократно

|  |  |
| --- | --- |
| Эмбеддинг запроса | [Текст] Эмбеддинг, для которого необходимо найти похожие строки в таблице. Значение для данного поля может быть получено с помощью блока Получить эмбеддинг. |
| База знаний | [Таблица Данных] Таблица, по которой будет осуществляться поиск. Одна из колонок таблицы должна содержать эмбеддинги, другие колонки таблицы могут содержать любые пользовательские данные. |
| ID базы | [Текст] Можно указать другой ID базы, чтобы блок при первом запуске начал строить базу данных векторов заново. При первом использовании блока происходит построение базы данных векторов и это может занимать длительное время. При последующих вызовах блока с одним и тем же "ID базы" будет использоваться уже созданная база данных векторов и скорость работы блока возрастает многократно |
| Колонка эмбеддингов | [Число/Текст] Индекс или название колонки из таблицы, указанной в свойстве "База знаний", в которой размещены эмбеддинги. По умолчанию 1. |
| Количество результатов | [Число] Максимальное количество возвращаемых блоком строк-результатов из базы знаний. По умолчанию 5. |
| Индексы строк | [Список] Список индексов строк из Базы знаний, содержащих наиболее похожие на запрос эмбеддинги, отсортированный в порядке убывания похожести. |
| Похожесть | [Список] Числовые меры похожести для каждого из результатов |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |

## Assistant Generator

Assistant Generator

|  |  |
| --- | --- |
|  | This Request |
|  |  |
|  | IsSimulate |
|  | Timeout |
|  | Response Selector |
|  | Response Type |
| Уровень обработки | Выбор уровня обработки ошибок. Возможные значения: "Default" - по умолчанию; "Ignore" - ошибки игнорируются; "Handle" - ошибки обрабатываются. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Уровень сообщений | Выбор уровня сообщений, которые будут выводить блоки при работе. Возможные значения: "Default" - по умолчанию; "Release" - вывод отключен; "Debug" - вывод основной информации; "Detailed" - вывод подробной информации. Если выбрано значение "Default", то будет использоваться значение блока "Старт" данной диаграммы |
| Текст ошибки | [Текст] Возвращает подробную информацию об ошибке в случае некорректного выполнения работы блока |