

Описание API ML компоненты

immediate

2 марта 2025 г.

Версия фреймворка 0.3.6

```
openapi: 3.1.0
info:
  title: ML Component - OpenAPI 3.1
  description: |-
    Спецификация сервиса MLComponent
  termsOfService: https://platform-dev-cs-hse.objectoriented.ru/terms/
  contact:
    email: apiteam@platform-dev-cs-hse.objectoriented.ru
  version: 1.0.0
#externalDocs:
#  description: Find out more about Swagger
#  url: http://swagger.io
servers:
  - url: http://localhost:8002
  - url: https://platform-dev-cs-hse.objectoriented.ru/app/mlcmp
tags:
  - name: predict
    description: Операции инференса
    externalDocs:
      description: Подробнее
      url: https://platform-dev-cs-hse.objectoriented.ru
  - name: status
    description: Операции статуса
security:
  - basic_auth: []
paths:
  /predict:
    post:
      tags:
```

```

- predict
summary: Выполнить инференс
description: |-
    Загружает модель и вызывает инференс функцию.

    В качестве аргументов передает входные данные запроса и загруженную модель.

    Возвращает выходы инференс функции.

    Если входные или выходные переменные имеют тип "FILE", то выполняет
    необходимые операции загрузки и выгрузки файлов в ящики.
operationId: predict
requestBody:
  description: Тело запроса
  content:
    application/json:
      schema:
        $ref: '#/components/schemas/PredictionRequest'
      examples:
        obj_in_obj_out:
          summary: 'Вход: объект; выход: объект'
          value:
            inputs:
              - name: X
                datatype: FP32
                shape: [2, 4]
                data: [1.0, 2.0, 6.0, 4.0, 1.0, 0.0, 1.0, 0.0]
            output_fields:
              - name: predict
                datatype: INT32
              - name: predict_proba
                datatype: FP32
      required: true
responses:
  '200':
    description: Операция завершена успешно
    content:
      application/json:
        schema:
          $ref: '#/components/schemas/PredictionResponse'
  '400':

```

```

description: Ошибка входных данных
content:
  application/json:
    schema:
      $ref: '#/components/schemas/ErrorResponse'
'401':
description: Аутентификация не пройдена
content:
  application/json:
    schema:
      $ref: '#/components/schemas/ErrorResponse'
'404':
description: Не найден заданный входной файл
content:
  application/json:
    schema:
      $ref: '#/components/schemas/ErrorResponse'
'422':
description: Ошибка при вызове инференс функции
content:
  application/json:
    schema:
      $ref: '#/components/schemas/ErrorResponse'
default:
description: Операция завершена успешно
content:
  application/json:
    schema:
      $ref: '#/components/schemas/PredictionResponse'
/modelversion:
get:
tags:
  - status
summary: Получить информацию о модели
description: |-
  Информация о модели и ее вызове
operationId: modelversion
responses:
'200':
description: Операция завершена успешно
content:

```

```

        application/json:
          schema:
            $ref: '#/components/schemas/ModelVersionResponse'
      default:
        description: Операция завершена успешно
        content:
          application/json:
            schema:
              $ref: '#/components/schemas/ModelVersionResponse'
/modelversion/{model_version}/license:
  parameters:
    - name: model_version
      in: path
      required: true
      schema:
        type: string
      description: |
        Версия модели
  get:
    tags:
      - status
    summary: Получить лицензию
    description: |-
      Получить лицензию, если лицензия поставляется сервисом
    operationId: modelversion_license
    responses:
      '200':
        description: Операция завершена успешно
        content:
          text/plain:
            schema:
              type: string
      '404':
        description: Файл лицензии не предоставляется сервисом
        content:
          application/json:
            schema:
              $ref: '#/components/schemas/ErrorResponse'
    default:
      description: Операция завершена успешно
      content:

```

```

        text/plain:
            schema:
                type: string
components:
    schemas:
        PredictionRequest:
            description: Объект с входами операции инференса
            type: object
            properties:
                model_parameters:
                    description: |
                        Параметры, передаются как аргумент в инференс функцию.
                        Каждый параметр представляется кортежем из двух элементов (имя, значение).
                    type: array
                    items:
                        type: array
                    prefixItems:
                        - type: string
                        - {}
                    minItems: 2
                    maxItems: 2
            inputs:
                type: array
                description: |
                    Входные переменные, передаются как аргумент в инференс функцию.
                items:
                    type: object
                    properties:
                        name:
                            description: Имя входной переменной.
                            type: string
                        data:
                            description: |
                                Значение входной переменной.
                                Если datatype == "FILE", то содержит строку с локацией файла.
                                Если datatype != "FILE", то содержит массив как сериализованную строку.
                            datatype:
                                description: Тип входной переменной.
                                type: string
                                enum: ["FP32", "FP64", "INT32", "FILE", "str", "dict"]
                        content_type:

```

```

description: |
    Тип содержимого файла, если тип входной переменной datatype == "FILE".
type: string
shape:
description: |
    Размерность, как массива, значения входной переменной.
    Применимо к любым типам данных входной переменной.
oneOf:
- type: array
  items:
    type: integer
  minItems: 1
  maxItems: 32
- type: integer
required: ["name", "data", "datatype", "shape"]
maxItems: 32
output_fields:
description: Список выходных переменных, которые инференс функция должна вернуть.
type: array
items:
  type: object
  properties:
    name:
      description: Имя выходной переменной.
      type: string
  data:
description: |
    Локация, куда следует сохранить выходной файл, если datatype == "FILE".

    Если datatype == "FILE", значение data соответствует файлу, то значение
    может быть сохранено в указанный файл.

    Если datatype == "FILE", значение data соответствует файловой группе, то
    значение может быть сохранено в созданный временный файл в заданной
    файловой группе.

    Если datatype == "FILE", но значение data не задано, то значение может быть
    сохранено в созданный временный файл в созданной временной файловой группе.

    Выражение "может быть" означает, что описанный сценарий будет выполнен,
    если инференс функция не подменит переданное ей в атрибуте data значение.

```

```

    type: string
  datatype:
    description: Ожидаемый тип выходной переменной
    type: string
    enum: ["FP32", "FP64", "INT32", "FILE", "str", "dict"]
  content_type:
    description: |
      Ожидаемый тип содержимого файла,
      если ожидаемый тип выходной переменной datatype == "FILE".
    type: string
  required: ["name", "datatype"]
  maxItems: 32
  required: ["inputs", "output_fields"]
PredictionResponse:
  description: Объект с выходами операции инференса
  type: object
  properties:
    outputs:
      description: |
        Выходные переменные, возвращаются как результат из инференс функции.
      type: array
      items:
        type: object
        properties:
          name:
            description: Имя выходной переменной.
            type: string
          data:
            description: |
              Значение выходной переменной.
              Если datatype == "FILE", то содержит строку с локацией файла.
              Если datatype != "FILE", то содержит значение как сериализованную строку.
          datatype:
            description: Тип выходной переменной.
            type: string
            enum: ["FP32", "FP64", "INT32", "FILE", "str", "dict"]
          content_type:
            description: |
              Тип содержимого файла, если тип выходной переменной datatype == "FILE".
            type: string
        shape:

```

```

description: |
    Размерность значения выходной переменной.
    Может быть массивом целых чисел или целым числом.
    Применимо к любым типам данных входной переменной.
oneOf:
  - type: array
    items:
      type: integer
    minItems: 1
    maxItems: 32
  - type: integer
    required: ["name", "data", "datatype", "shape"]
maxItems: 32
ModelVersionResponse:
description: Метаданные модели и информация о использовании
type: object
properties:
  model_version:
    description: Версия модели
    type: string
  license:
    description: URL для получения лицензии
    type: string
  openapi_spec:
    description: Спецификация API для вызова модели в формате OpenAPI JSON
    type: object
required: ["model_version", "license", "openapi_spec"]
ErrorResponse:
description: Объект с результатом операции инференса, которая не завершилась успешно
type: object
properties:
  errors:
    description: Список ошибок.
    type: array
    items:
      description: Ошибка при выполнении операции инференса.
      type: object
      properties:
        type:
          description: Тип ошибки.
          type: string

```



```
    message:
      description: Сообщение с описанием ошибки.
      type: string
      required: ["type", "message"]
    maxItems: 10
securitySchemes:
  basic_auth:
    description: Базовая аутентификация
    type: http
    scheme: basic
```