**ML Train & Forecast**

**구동 규격서**

(Flask framework base.)

SP Technology

**Document History**

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Description | Name | Date |
| 0.1.0 | 초안 작성 | 박남규 | 2021-07-27 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**목차**

[1. ML Train & Forecast 5](#_Toc78360369)

[1.1. 구동 설정 5](#_Toc78360370)

[1.1.1. 모델 생성 매니페스트 설정 5](#_Toc78360371)

[1.1.2. 데이터베이스 설정 7](#_Toc78360372)

[1.1.3. Repository 설정 9](#_Toc78360373)

# ML Train & Forecast

## 구동 설정

Flask 기반 ML 학습 및 예측 서비스 실행하기 위한 구동은 다음과 같이 설정할 수 있습니다. 현재 Flask 기반 ML 학습 기능은 시계열 데이터를 학습하여 예측하는 기능만 제공합니다.

### 모델 생성 매니페스트 설정

모델 학습에서 필요한 데이터 수집 방법에서부터 전처리 방법, 알고리즘, 샘플링, 하이퍼 파라미터 설정 및 생성된 모델을 저장할 레파지토리를 위치 등을 설정해야 합니다.

**옵션 설명**

|  |  |
| --- | --- |
| **필드명** | **설명** |
| name | 설정한 이름은 서비스 및 관리에 사용하게 됩니다. |
| database | 가져올 데이터베이스 이름을 지정합니다.  동일할 서비스 영역에 데이터베이스 설정 이름을 지정합니다. |
| filter | 데이터 셋을 구성하기 위한 필터 정보를 지정합니다.  동일 서비스 영역의 ***filters***에 기술된 필터 이름을 지정합니다. |
| period | 생성될 모델의 예측 기간을 지정합니다.  동일 서비스 영역의 ***periods***에 기술된 period 이름을 지정합니다. |
| preprocessing | 모델 학습 및 예측에서 사용하는 데이터는 ML 알고리즘에 알맞은 데이터로 바꿔주는 과정이 필요합니다.  동일 서비스 영역의 ***preprocessing***에 기술된 전저처리 이름을 지정합니다. |
| feature | 모델 학습 및 예측에서 사용할 피처를 지정해야 합니다. 데이터베이스에서는 Column이 될 수 있습니다.  동일 서비스 영역의 ***features***에 기술된 feature 이름을 지정합니다. |
| generator | 모델 학습 및 예측에서 기존 데이터에는 없지만 새로운 데이터가 필요할 수 있습니다.  동일 서비스 영역의 ***generators***에 기술도리 generator 이름을 지정합니다. |
| algorithm | 생성될 모델의 알고리즘을 지정해 줘야 합니다.  \* 제공하는 모델 알고리즘   * LSTM * Bi-LSTM |
| hyperparametor | 하이퍼 파라미터 설정을 모델 학습시에 사용합니다.  동일 서비스 영역의 ***hyperparameter***에 기술된 hyperparametor 이름을 지정합니다. |
| repository | 모델 생성 후 저장될 레파지토리를 지정해야 합니다.  동일 서비스 영역에 ***repository***에 기술된 repository 이름을 지정합니다. |

**모델 생성 설정하기**

|  |
| --- |
| exports:  - name: cmp-ai-host  database: cmp-rsc-elasticsearch  filter: cmp-host-filter  period: [cmp-period-v1, cmp-period-v2]  preprocessing: cmp-preprocessing  feature: cmp-feature  generator: cmp-generator  algorithm: bi-lstm  hyperparametor: None  repository: cmp-model-repos |

### 데이터베이스 설정

학습에 사용할 데이터베이스를 설정하는 옵션입니다.

**옵션 설명**

|  |  |
| --- | --- |
| **필드명** | **설명** |
| name | 데이터베이스 정보를 참조할 때 사용할 이름을 설정합니다. |
| driver | 데이터베이스 사용하는 드라이버를 설정합니다.  \* 드라이브 종류   * csv * mongodb * elasticsearch |
| uri | 데이터베이스 접속 URI를 설정합니다.  \* URI 형식   * CSV 참조:   *“*[*file:///root/data/temp\_20210621.csv*](file:///root/data/temp_20210621.csv)*”*   * Mongo 참조:   *"mongodb://root:1234@localhost:7017/cmp-ai?authSource=admin&authMechanism=SCRAM-SHA-1”*   * Elasticsearch 참조:   [*http://210.217.178.154:9200*](http://210.217.178.154:9200) |
| collection | MongoDB 사용시 값을 설정합니다. 사용할 collection을 설정합니다. |
| index | Elasticsearch 사용시 값을 설정합니다. 사용할 index를 설정합니다. |
| doc | Elasticsearch 사용시 값을 설정합니다. 사용할 document를 설정합니다. |

**CSV 파일 사용하기**

|  |
| --- |
| database:      - name: cmp-rsc-csv        driver: csv        uri: 'file:///root/git/cmpai-v2-train/data/stat\_host\_graph\_temp\_20210621.csv' |

**MongDB 사용하기**

|  |
| --- |
| database:    - name: cmp-rsc-mongodb      driver: mongodb      uri: "mongodb://root:1234@localhost:7017/cmp-ai?authSource=admin&authMechanism=SCRAM-SHA-1"      collection: cmp-ai-resource |

**Elasticsearch 사용하기**

|  |
| --- |
| database:    - name: cmp-rsc-elasticsearch        driver: elasticsearch        uri: 'http://210.217.178.154:9200'        index: cmp-ai-resource        doc: cmp-ai |

### Repository 설정

학습 후 생성된 모델을 저장할 레파지토리(Repository)를 설정합니다.

**옵션 설명**

|  |  |
| --- | --- |
| **필드명** | **설명** |
| name | 사용할 레파지토리(Repository) 정보를 이름을 설정합니다. |
| driver | 레파지토리에 접속시 사용하는 드라이버를 설정합니다.  \* 드라이브 종류   * nas : 로컬 시스템의 스토리지에 접속 |
| uri | 레파지토리 접속 URI를 설정합니다.  \* URI 형식   * NAS:   *“*[*file:///root/repository/model*](file:///root/repository/model)*”* |

**로컬 시스템 저장소 사용하기**

|  |
| --- |
| repository:  - name: cmp-model-repos  driver: nas  uri: 'file:///root/git/cmpai-v2-train/model' |