

Release Notes

1. 현 페이지는 엔지니어가 수기로 작성하였습니다. 따라서 최신화를 보장하지 않습니다.
2. v1.7.0 부터 패키지 네이밍 규칙이 변경되었습니다.

- v1.8.1 - 27 Sep 2023
- v1.8.0 - XX Sep 2023 (날짜미정)
- v1.7.0 - 13 Jul 2023
- v1.6.2.hotfix1 - 2 Jun 2023
- v1.6.2 - 17 Mar 2023
- v1.6.0.rc2 - 5 Dec 2022
- v1.6.0.rc1 - 05 Sep 2022
- v1.5.8.rc5
- v1.5.8.rc4 - 10 Jun 2022
- v1.5.8.rc3 - 12 May 2022
- v1.5.8.rc2 - 29 Nov 2021
- v1.5.8.rc1 - 29 Nov 2021
- Package Naming
- 학습데이터

v1.8.1 - 27 Sep 2023

Bug Fixes and Other Changes

- [AI-3247][bugfix] Remove 'n_estimator' parameter from ExtraTreeClassifier
- [AI-3248][bugfix] Fix predicted label when loss function of catboost model is LogLoss

PACKAGING_VERSIONS

- ml_install-release-1.8.1-167.tar.gz
- ml_infra: release/1.8.1 - 1d69027039d87c28cae47e59487227e6c30eb6c6 - 2023-09-27T12:59:15.637+09:00
- ml_train: release/1.8.1 - 89f3be4c5cafec48d14b3e5b6449c28d3a0d60cf - 2023-09-27T13:00:32.403+09:00
- ml_pred: release/1.8.1 - 7f05b0de2a60eaca4d216b6fca78eed9ec794a19 - 2023-09-27T13:00:06.104+09:00
- ml_management_api: release/1.8.1 - 1a7a58278c5c0d9aadbeef973f04cab38b6190a1 - 2023-09-27T12:58:29.360+09:00

v1.8.0 - XX Sep 2023 (날짜미정)

Major Features and Improvements

- [AI-3236][feature] application 배포 방식 변경: definitionfile file → Helm Chart
- [AI-3177][feature] 플랫폼 모니터링 API 개발: Implementing collect prediction monitoring data
- [AI-3231][feature] Syncer 탐지 속도 개선 370sec → 3 sec

Bug Fixes and Other Changes

- [AI-3238][feature] 클러스터 설치 값 옵션화
 - Set NFS server, Change NodePort Range, Enable Addon
- [AI-3222][bugfix] Fix wrong hyperparameter style for GradientBoostingClassifier

- [AI-3228][bugfix] IsolationForest exponential KDE로 fake data 생성 시 NotImplementedError 버그 발생. 해당 커널 지원하지 않는 것으로 변경
- [AI-3230][bugfix] IsolationForest root_fix 학습 시 에러 발생
- [AI-3229][feature] 예측 주기 기본 값 변경
 - max_buffer_size: 524288 → 131072
 - max_latency: 60 → 10
 - kafka.timeout: 10 → 5
- [AI-3234][bugfix] global.properties standalone trainer 설정 버그픽스
- [AI-3245][bugfix] CatBoostClassifier BinaryClassification task lossfunction 과 target value 미스매치 버그
 - 이전에는 다중분류만 지원하였으나 이진분류도 지원함으로써 타겟변수의 포맷에 변경 필요해짐. 버그 픽스 완료

PACKAGING_VERSIONS

- ml_install-release-1.8.0-166.tar.gz
- ml_infra: release/1.8.0 - 1d69027039d87c28cae47e59487227e6c30eb6c6 - 2023-09-04T13:54:49.833+09:00
- ml_train: release/1.8.0 - a4359e336948ea3ce2f4417bd41605bf136740f9 - 2023-09-04T13:55:13.240+09:00
- ml_pred: release/1.8.0 - 6da4588bb3922a24bf801d6689d126de74be3339 - 2023-08-31T15:29:14.843+09:00
- ml_management_api: release/1.8.0 - 1a7a58278c5c0d9aadbeef973f04cab38b6190a1 - 2023-08-29T12:35:26.873+09:00

v1.7.0 - 13 Jul 2023

v1.7.0 부터 패키지 네이밍 규칙이 변경되었습니다.

Bug Fixes and Other Changes

- [AI-3169][bugfix] 비지도 모델을 paruqet 파일로 학습 시 모듈 에러 발생, 다중 트레이너의 경우 로그 중복 현상 발생, Multi-Anomaly-Detection 타입 예측 에러 발생
- [AI-3186][bugfix] 패키지 설치 시 docker 설치 부분 빠짐
- [AI-3194] StackingClassifier의 주요 피처가 text 타입만 있을 경우, XAI 예측을 수행하지 않도록 수정
 - 기존에는 에러를 발생시키고 피처를 변경 ex)squeeze=False 로 해결하였으나, 해당 버전부터는 XAI 예측을 하지않도록 수정
- [AI-3195] 모든 모듈 버전을 패키지 버전과 동일하도록 변경
- [AI-3193] 모든 오브젝트에 label과 annotation 설정
- [AI-3170] 예측 엔진 설정 시 기존 예측 엔진 값 유지
- [AI-3192] 초기 엔진 설정 및 예측 엔진 설정 창의 각 파라미터 description 한국어로 변경

PACKAGING_VERSIONS

- ml_install-release-1.7.0-161.tar.gz (2023.07.12)
- ml_infra: release/1.7.0 - 14cfa63f5a751a532053b0d0b3ab267e99aed6f - 2023-07-12T14:51:00.988+09:00
- ml_train: release/1.7.0 - 74187e261090aef2b8a1ef3910e236cad3f2f6a3 - 2023-07-12T13:56:25.497+09:00
- ml_pred: release/1.7.0 - c464724d0c84ee9e6655357b20a9d71e73f40f62 - 2023-07-05T16:58:34.919+09:00
- ml_management_api: release/1.7.0 - 1a7a58278c5c0d9aadbeef973f04cab38b6190a1 - 2023-07-05T17:00:36.608+09:00

v1.6.2.hotfix1 - 2 Jun 2023

Jun 2 는 ml-hotfix 발행일로 전체 AI 패키지 핫픽스 발행은 아직 하지 않음

Known Issues

- 비지도 모델을 paruqet 파일로 학습 시 모듈 에러 발생
 - parquet 대신 pyarrow를 사용하기 때문
- 다중 트레이너의 경우 로그 중복 현상 발생
- Multi-Anomaly-Detection 타입 예측 에러 발생
- 패키지 설치 시 docker 설치 부분 빠짐

Bug Fixes and Other Changes

- 예측 시 node_selector, cpu 필드 설정 버그 픽스

- TextClassifier 의 uncertainty 점수 버그 픽스
- 리눅스 헤더 설치 디렉토리 경로 \$HOME/ml_patch 오타 버그 픽스
- GPU 드라이버 설치 안 함 옵션 추가
- 클러스터 환경에서 GPU 드라이버 업로드 횟수 줄임
 - 각 노드 별 N번 업로드 → 설치 노드에만 한 번 업로드
- 레지스트리 빌드 시 도커 이미지 검증 단계 추가

PACKAGING_VERSIONS

- ml_install-hotfix-1.6.2.hotfix1-158.tar.gz
- ml_infra: hotfix/1.6.2.hotfix1 - 19f5da10f026b139a0626ba2fc731dfc9da9f1cb - 2023-06-02T14:41:31.980+09:00
- ml_train: hotfix/1.6.2.hotfix1 - aee795ccb65adbb98aa1c10abbf7560d99d07023 - 2023-06-02T14:14:30.361+09:00
- ml_pred: hotfix/1.6.2.hotfix1 - 84f25b8fc6ebfd3cb6106290bc4d1a034c4a70e9 - 2023-06-02T14:17:36.970+09:00
- ml_management_api: hotfix/1.6.2.hotfix1 - ea43a5ce52a79490978d0a181a378aace0654e8a - 2023-06-02T14:18:45.374+09:00

v1.6.2 - 17 Mar 2023

Major Features and Improvements

- UI에서 학습과 예측 설정 가능

예측 설정

engine

Node Selector: kubernetes.io/hostname [value] - +
Reset to Default

GPU_device_id: null
Reset to Default

CPU: 0.3
Reset to Default

N_engines: 2
Reset to Default

prediction_model_params

batch size: 4096
Reset to Default

prediction_time

Max_buffer_size: 52428
Reset to Default

Max_latency: 60
Reset to Default

취소 예측 시작

- 학습 알고리즘 Unit Test 기능 추가

Bug Fixes and Other Changes

- 예측 파일 저장 시 온점(.)이나 ; 빈칸이 포함된 열은 삭제
 - 처리 이유: 수집과 전처리에서 중복 필드가 들어오면 pyarrow저장 시 payload, payload.1 변형됨. 이를 미니언즈에서 읽을 때 에러 발생

- pyarrow version update v10.0.1
 - backend에서 버전 업데이트하여 기존 버전으로는 parquet 파일 읽기 불가
- 클러스터 노드 추가/제거 에러 발생 시 무기한 대기 이슈 해결
- GPU 드라이버 경로 오타 수정: \$HOME/ml_patch
- TextCatBoostClassifier shrink시 텍스트 피쳐 없을 경우 에러 발생 픽스
- 플랫폼 매니저에서 예측 데이터(토픽) 리스트 불가 이슈 픽스

Known Issues

- 예측 시 node_selector, cpu 필드 설정 불가 이슈
- TextCatBoostClassifier 의 uncertainty 점수 버그 발생
- 리눅스 헤더 설치 디렉토리 경로 \$HOME/ml_patch 오타 버그

PACKAGING_VERSIONS

- ai-1.6.2.rc3-3-pkg.tar.gz
- prep_app: release/1.6.2.rc3 - 25fddd593cadfb44d9a430059e3a274b6713ce5c - 2023-03-08T21:11:38.766+09:00
- api-gateway: release/1.6.2.rc3 - 25fddd593cadfb44d9a430059e3a274b6713ce5c - 2023-03-08T21:11:38.766+09:00
- collector-app: release/1.6.2.rc3 - 25fddd593cadfb44d9a430059e3a274b6713ce5c - 2023-03-08T21:11:38.766+09:00
- discovery: release/1.6.2.rc3 - 25fddd593cadfb44d9a430059e3a274b6713ce5c - 2023-03-08T21:11:38.766+09:00
- prep-service: release/1.6.2.rc3 - 25fddd593cadfb44d9a430059e3a274b6713ce5c - 2023-03-08T21:11:38.766+09:00
- resource-monitor: release/1.6.2.rc3 - 25fddd593cadfb44d9a430059e3a274b6713ce5c - 2023-03-08T21:11:38.766+09:00
- zookeeper-app: release/1.6.2.rc3 - 25fddd593cadfb44d9a430059e3a274b6713ce5c - 2023-03-08T21:11:38.766+09:00
- ml_infra: release/1.6.2.rc3 - a3806c321e9e32973becba26553302a5dc9aec8f - 2023-03-08T16:52:12.906+09:00
- ml_train: release/1.6.2.rc3 - f539680158da86c33f4f075b5b3a6d5f7fd64b8c - 2023-03-08T16:53:08.589+09:00
- ml_pred: release/1.6.2.rc3 - 2da3c8b0dc4f089fb46c84971ef99b64ecd91887 - 2023-03-08T16:53:46.952+09:00
- ml_management_api: release/1.6.2.rc3 - ea43a5ce52a79490978d0a181a378aace0654e8a - 2023-03-08T16:54:08.852+09:00
- nox: release/1.6.2.rc3 - 79d8e8a60d9e6e8edadfd1bb012e3d8ac3d8615a - 2023-03-08T17:21:48.555+09:00
- minions: release/1.6.2.rc3 - a4ea9be983235fc6cb045a8431fd7a4861bfb1a6 - 2023-03-08T17:22:15.681+09:00
- lucy:
- manage: release/1.6.2.rc3 - 3dc60c9203a51b0cef08c123a8c2d246ef78dfe2 - 2023-03-08T17:30:22.937+09:00

v1.6.0.rc2 - 5 Dec 2022

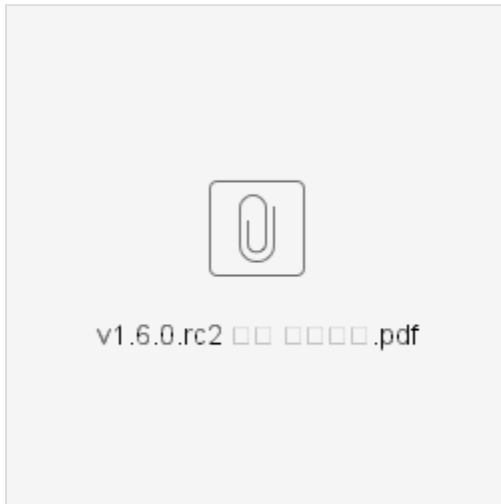
Bug Fixes and Other Changes

- Hostname 에 '-' 문자 허용
- IsolationForest 알고리즘 학습 시, 모든 피쳐가 범주형 또는 연속형일 경우, scatter 생성 에러 픽스
- TextCatBoostClassifier 알고리즘 shrink 학습 시 연속형 변수만 있을 경우 발생하는 에러 픽스
- TextCatBoostClassifier 알고리즘의 loss_function 파라미터 UI 선택 가능
- NFS 중복 설치 시도 시 설치 중단
 - v1.6.0.rc1 에서 발생한 'NFS 데몬이 로드되지않거나 비정상일 경우, 설치 확인이 제대로 안되어 중복 설치하는 이슈' 의 해결책

Known Issues

- Allocator attributeError 발생 → 문서 작성하여 전달

- 엔진 에러로 예측 한 번 수행 후 엔진 종료 → 문서 작성하여 전달



- 비지도학습 시 피쳐가 3개인 경우, Scatter 계산 에러 발생 → script 작성하여 전달 (위 2가지 이슈도 함께 포함)

```
jintegjinte:~/hotfix$ python3 hotfix.py
[INFO] Copy iforest_new.py -> /ai/gauss_ml/ml_train/training_framework/tframe/algorithms/iforest_new.py
[INFO] Copy minibatch_kmeans.py -> /ai/gauss_ml/ml_train/training_framework/tframe/algorithms/minibatch_kmeans.py
[INFO] Copy processor.py -> /ai/gauss_ml/ml_pred/gauss_pipe/processor.py
[INFO] Copy run_allocator.py -> /ai/gauss_ml/ml_infra/kubernetes/gpu/run_allocator.py

=====
Now you can build image. Please manually run the command below.

1. /ai/gauss_ml/ml_infra$ make
2. /ai/gauss_ml/ml_infra/kubernetes$ make deploy
3. /ai/gauss_ml/ml_infra/kubernetes$ python3 gauss-cluster.py restart
```

PACKAGING_VERSIONS

- ml_install-release-1.6.0.rc2-147.tar.gz674df798a883664a29dc15cbec1b6726bda08f2d - 2023-02-21T13:23:39.063+09:00
- ml_infra: release/1.6.0.rc2 - 609618b080b6eb426eab07821b54fde25018e1e9 - 2022-12-02T17:38:12.234+09:00
- ml_train: release/1.6.0.rc2 - 7d79abafa16b12814c8c0a3cdbab7784c0fed4a8 - 2022-12-05T08:46:53.957+09:00
- ml_pred: release/1.6.0.rc2 - 59d728f229ba378fb2092626086363a0510d7ef7 - 2022-11-21T21:00:48.449+09:00
- ml_management_api: release/1.6.0.rc2 - 7054e9aae0ea538e07c9a51555753f6405908866 - 2022-11-21T21:01:40.609+09:00

v1.6.0.rc1 - 05 Sep 2022

Breaking Changes

- 오케스트레이터를 docker swarm 에서 Kubernetes 로 변경
 - docker 커맨드 대신 kubectl 커맨드 사용
- ML 로그 저장 시 NFS 서버 사용
 - ML의 모든 노드에 저장하던 데이터를 하나의 NFS 서버에만 저장
- ML 설정 파일 단순화
 - engine_resources.yml, training_resources.yml, global.properties 3개의 파일을 global.properties로 통합

Major Features and Improvements

- 지도 학습 신규 알고리즘 6종 추가

- 'DecisionTreeClassifier', 'ExtraTreeClassifier', 'RandomForestClassifier', 'ExtraTreesClassifier', 'AdaBoostingClassifier', 'GradientBoostingClassifier'
- 비지도 학습 신규 알고리즘 MiniBatchKMeans 추가
- CatBoostClassifier가 텍스트 필드를 지원하는 TextCatBoostClassifier 알고리즘으로 업그레이드
- TextCatBoostClassifier 알고리즘 XAI 지원
- 예측 엔진 설정 UI 지원
- StackingClassifier 알고리즘에 max_len_limit 파라미터 설정 지원
- default 값 500에서 900으로 변경. UI에서 파라미터 설정 가능
- 예측 데이터 저장 시 parquet 포맷 지원
- GPU 디바이스 아이디 검증 기능 지원
- 레지스트리 이미지 삭제 기능 지원
- 삭제 시 최신 이미지 5개를 제외한 모든 이미지 삭제
- 쿠버네티스 대시보드 추가
- 로그, 자원 사용량 모니터링 가능
- catboost 버전 업데이트: v0.22 --> v1.0.3
- 예측 파라미터
 - 최대 예측 데이터 갯수 기본값 변경: 1,048,576 --> 524,288
 - 최대 예측 데이터 갯수 설정 경로 변경: ml.pred.max_buffer_size --> ml.pred.kafka.max_buffer_size
 - 엔진 리커버리 타임 기본값 변경: 30분 --> 10분
 - 엔진 리커버리 타임 설정 지원: ml.pred.kafka.recovery_time
 - 최대 데이터 대기 시간 기본값 변경: 45분 --> 30분
 - 최대 데이터 대기 시간 설정 지원: ml.pred.kafka.max_nodata_time
- model_api.start_training
 - 파라미터명 변경: constraints --> node_selector
 - trainers, bypass_allocation 파라미터 삭제. 학습 노드 선택할 경우 node_selector 파라미터 사용
- engine_api.format_pred_options
 - 예측 노드 선택을 위한 hostname 파라미터 추가
 - 예측 GPU 사용을 위한 gpu_device_id 파라미터 추가
 - 예측 노드 자원 사용량 제어를 위한 cpu, memory 파라미터 추가
- pandas non-utf8 인코딩 형식 지원
 - non-utf8 하둡 데이터 읽기 지원 (단, 한글의 경우 바이너리로 읽힐 수 있음)

Bug Fixes and Other Changes

- 환경변수 이름 변경
 - W\$DV --> \$PV
 - W\$DS --> \$KUBE
- xai_framework
 - 모델의 피처가 하나일 경우, XAI 에러 발생하는 이슈 수정
- Tokenizer 모듈 변경
 - tensorflow.keras.preprocessing.text.Tokenizer 에서 자체 Tokenizer로 변경
- training_framework 에서 tframe으로 모듈 이름 변경
- 노드 선정 단계를 k8s에 일임
 - 복잡한 finite, infinite 조건과 노드 점수 계산 방식 삭제
 - k8s의 노드 필터링과 노드 스코어링 사용
- 패키지 설치 시 개별 리눅스 헤더 지원 가능
 - 현재 패키지에서 지원하는 리눅스 헤더(5.4.0-65, 5.4.0-91, 5.8.0-43)와 필요한 버전이 있을 시, \$HOME/ml_patch 에 엔지니어가 추가
- 패키지 설치 시 개별 GPU 드라이버 지원 가능
 - 현재 패키지에서 지원하는 드라이버 버전 (450.142, 460.80) 외 필요한 버전이 있을 시, \$HOME/ml_patch 에 엔지니어가 추가

Deprecations

- engine_resources.yml, training_resources.yml 사용 안함
- ml_utils 모듈을 ml_train으로 통합
- cpu 도커이미지 생성 하지 않음

Known Issues

- Hostname 에 '-' 있을 경우, 에러 발생
- IsolationForest 알고리즘에서 모든 피처가 범주형 또는 연속형일 경우, scatter 생성 에러 발생
- NFS 데몬이 로드되지않거나 비정상일 경우, 설치 확인이 제대로 동작하지 않아 중복 설치하는 이슈 발생

PACKAGING_VERSIONS

- ai-v1.6.0.rc1-7-pkg.tar.gz, ml_install-release-1.6.0.rc1-139.tar.gz
- Build number: #139
- ml_infra: release/1.6.0.rc1 - 2ed083172e5afd7b17b565a646bb9a8a2a747a6b - 2022-09-05T17:54:40.915+09:00
- ml_train: release/1.6.0.rc1 - af15ceee541f8f1223cee271800e5eaf4bc50bc4 - 2022-09-01T19:03:45.803+09:00
- ml_pred: release/1.6.0.rc1 - 59d728f229ba378fb2092626086363a0510d7ef7 - 2022-09-02T20:04:02.182+09:00
- ml_management_api: release/1.6.0.rc1 - 7054e9aae0ea538e07c9a51555753f6405908866 - 2022-09-01T18:05:30.854+09:00

v1.5.8.rc5

1.5.8.rc4 와 동일

v1.5.8.rc4 - 10 Jun 2022

Major Features and Improvements

- stacking classifier XAI 기능 추가됨.

Known Issues

- engine_resources.yml과 training_resources.yml 생성 시 Cluster모드일 경우, 포트 기입란이 빠져있음. ---> 기본 포트 22 아닐 경우, deploy 에러 발생
 - how to fix it

edit init_resources.py

```
# master:$DS/init_resources.py
# Add 128 line

126     else:
127         logger.info("Cluster mode")
128         base["ssh_port"] = ssh_port
129
130         if (ml_server is None) or ml_server.lower() ==
"localhost":
131             logger.error("No ML_SERVER in global.properties
despite it is CLUSTER MODE. Modify if first.")
132             exit(1)
```

PACKAGING_VERSIONS

- Build number: #125
- ml_infra: release/1.5.8.rc4 - dc94ae9aa6901a23d3627a04c124a1fe7b78678d - 2022-06-10T10:53:01.443+09:00
- ml_train: release/1.5.8.rc4 - 546f81a2cae27863340fb5ceadd3b05681488810 - 2022-06-10T11:00:55.354+09:00
- ml_pred: release/1.5.8.rc4 - 8d0fd8549ccae721cadf809a3f9f6ea8528d3f55 - 2022-06-10T10:59:48.370+09:00
- ml_utils: release/1.5.8.rc4 - 1e8799c8eccc9a95811796d9c4cc843b2739b277 - 2022-06-10T11:05:16.261+09:00
- ml_management_api: release/1.5.8.rc4 - c1b4ddb14a948cc6bd2ea6a3411c249c5c762dae - 2022-06-10T10:57:26.744+09:00

v1.5.8.rc3 - 12 May 2022

Bug Fixes and Other Changes

- /api/ml/key_filters/get API 요청 시 데이터가 없는 경우 리턴값 변경
 - empty array도 JSON 포맷으로 ml파트에서는 상관없으나, 백엔드에서 data 값을 파싱하여 사용하므로 {} 로 변경
- ```
{"data": [], "status": 200, "message": "ok"} --> {"data": {}, "status": 200, "message": "ok"}
```

## PACKAGING\_VERSIONS

- Build number: #124
- ml\_infra: release/1.5.8.rc3 - dc94ae9aa6901a23d3627a04c124a1fe7b78678d - 2022-05-12T15:52:53.031+09:00
- ml\_train: release/1.5.8.rc3 - 546f81a2cae27863340fb5ceadd3b05681488810 - 2022-05-12T15:38:07.188+09:00
- ml\_pred: release/1.5.8.rc3 - 8d0fd8549ccae721cadf809a3f9f6ea8528d3f55 - 2022-05-12T15:44:16.839+09:00
- ml\_utils: release/1.5.8.rc3 - 1e8799c8eccc9a95811796d9c4cc843b2739b277 - 2022-05-12T15:52:32.334+09:00
- ml\_management\_api: release/1.5.8.rc3 - c1b4ddb14a948cc6bd2ea6a3411c249c5c762dae - 2022-05-12T15:41:37.594+09:00

## v1.5.8.rc2 - 29 Nov 2021

---

## PACKAGING\_VERSIONS

- Build number: #123
- ml\_infra: release/1.5.8.rc2 - c781f5e26486a83b145d784127782ab2f4a68690 - 2021-11-29T13:45:18.303+09:00
- ml\_train: release/1.5.8.rc2 - deaa06f9f4d36487ccc24483c03c57285fbfbaf - 2021-11-29T13:53:56.448+09:00
- ml\_pred: release/1.5.8.rc2 - 8d0fd8549ccae721cadf809a3f9f6ea8528d3f55 - 2021-11-29T13:52:10.476+09:00
- ml\_utils: release/1.5.8.rc2 - 1e8799c8eccc9a95811796d9c4cc843b2739b277 - 2021-11-29T13:55:06.030+09:00
- ml\_management\_api: release/1.5.8.rc2 - c1b4ddb14a948cc6bd2ea6a3411c249c5c762dae - 2021-11-29T13:50:57.257+09:00

## v1.5.8.rc1 - 29 Nov 2021

---

### ml\_infra

- run\_training 리팩토링
- 이름 변경: run\_training.py --> run\_trainer.py
- 이에 따라 gen\_stack\_file 과 init\_resources.py 도 수정
- 파이썬 실행파일 모두 .py 확장자 같게끔 변경
  - gauss-swarm --> gauss-swarm.py
  - gen\_stack\_file --> gen\_stack\_file.py
  - init\_global\_prop --> init\_global\_prop.py
  - Makefile 내의 init\_global\_prop 도 수정: init\_resources --> init\_resources.py
- gen\_stack\_compose.py 에 주피터 볼륨 생성 로직 추가



- gpu/pygloo 에서 pygloo가 정상적으로 실행되지 않던 이슈 수정
- init\_resources.py 에서 node resources 가 중복되던 현상 수정
- docker build 가 이제 BuildKit 을 사용합니다. 빌드 속도가 약 10~20% 정도 더 빨라졌습니다. 압축 시 용량이 더 줄어 듭니다.

## ml\_train

### fandas

- `fandas` 가 이제 독립 모듈로 작동합니다.
- `fandas.read_hdfs` 가 학습 데이터를 캐싱 합니다. 타겟 데이터는 `/tmp/<hash>` 로 저장되며 이 해시값은 파일의 사이즈와 파일의 첫 1MB 와 마지막 1MB 를 종합적으로 해시하여 저장합니다. 다운로드 된 파일은 기본적으로 30일 뒤 자동으로 삭제 됩니다
- `fandas.hdfs_checksum` : 하둡에있는 파일을 해싱합니다
- `fandas.local_checksum`: 로컬에 있는 파일을 해싱합니다.
- `fandas.receipts.unix.delayed_delete`: 로컬에 있는 파일을 원하는 시간 이후에 삭제합니다. (패키지 `at` 이 깔려있어야 합니다.)
- `cache==True` 일 때는 `csv`와 `parquet` 모두 알아서 읽을 수 있습니다.
- `fandas.read_parquet` 가 지원 됩니다. `pyarrow` 또는 `fastparquet` 가 있어야 합니다. (없으면 함수 자체가 생성되지 않습니다.)
- `fandas.read_any` 가 지원 됩니다. 자동으로 `parquet` 와 `csv` 를 구분합니다.

### tframe

- 안정화 전까지는 `training_framework` 와 동일하게 병렬로 업데이트 될 예정입니다.
- `training_framework` 를 대신하는 새로운 모듈입니다.
- `run_trainer` 가 `init` 에 정의됩니다 -> `(from tframe import run_trainer)`
- legacy 코드가 삭제되었습니다.
- log 메시지가 더 많은 정보를 보여줍니다.

### Bug Fixes and Other Changes

- `mltk`, `mmh3` 및 라이브러리를 활용 하는 부분이, python 빌트인 함수로 대체

### Deprecations

- 예측 시 `kafka reader`의 `Kafka poll time` 로그 삭제

## Package Naming

---

- v1.7.0 부터 Semantic Versioning 표기 방식을 따르도록 한다. (2023.7)

4.2.1

MAJOR Minor patch

- MAJOR: 기존 버전과 호환되지 않는 변경
- MINOR: 기존 버전과 호환되면서 새로운 기능을 추가
- PATCH: 기존 버전과 호환되면서 버그를 수정
- 더 자세한 사항은 Semantic Versioning 2.0.0 문서를 참고한다. <https://semver.org/>
- hotfix 버전은 배포버전 뒤 hotfixN으로 표기하도록 정함 (2023.5)
- 끝자리 올수는 개발 버전을, 찍수는 배포 버전을 의미

- Fix version 설정 방법

1. rc1에서 이슈 발생 후 rc2에서 Done, Duplicate로 보완 될 경우 Fix Version/s : rc2로 설정



2. rc1에서 이슈 발생 후 rc3에서 Done, Duplicate로 보완 될 경우 Fix Version/s : rc3으로 설정



3. rc1에서 이슈 발생 후 해당 이슈가 Decline, Won't Do, Cannot Reproduce 경우 Fix Version/s : rc1으로 설정



4. rc1에서 이슈 발생 후 이번 Release에서 보완되지 못하고 In Code Review, Unresolved, Open, Reopen 상태로, 차기 Release에서 보완될 경우 Fix Version/s : Next Release 버전(품질보증팀에서 관리)으로 설정



## 학습데이터

- 학습데이터는 개발팀이 아닌 QA팀 서버에서 관리
- 테스트를 위해 QA팀에 문의하여 다운로드 받음
- 본 데이터는 2023년 1월 20일 QA 서버에 업로드된 데이터로 지도학습 label은 normal 과 anomalies로 구성

