### rforest

랜덤 포레스트 모델(Random Forest, 여러 개의 결정 트리를 임의로 학습하는 방식)을 사용하여 예측 데이터를 출력합니다.

#### 문법

사전에 준비된 학습 모델에 기반한 예측

rforest [size=INT] model=MODEL

서브쿼리를 통해 랜덤 포레스트 분석을 실행

rforest [size=INT] target=TARGET\_FILED FIELD, ... [ SUBQUERY ]

필수 매개변수

**FIELD, ...**

랜덤 포레스트 분석에서 사용할 필드들을 입력합니다.

**model=MODEL**

랜덤 포레스트 모델의 이름을 지정합니다. 로그프레소 엔진에 CLI로 접속하여 학습 모델을 생성하고 학습을 진행할 수 있습니다.

**target=TARGET\_FIELD**

랜덤 포레스트 분석에서 target 변수에 분류값으로 사용할 필드를 지정합니다.

**[ SUBQUERY ]**

랜덤 포레스트 분석에서 사용할 트레이닝셋 데이터 조회 쿼리를 입력합니다.

선택 매개변수

**size=INT**

포레스트를 구성하는 트리 갯수(기본값: 100)

#### 설명

이 명령어를 실행하면 **\_guess** 필드에 target 변수의 값을 추정하여 출력합니다.

#### 사용 예

이름이 rforest\_titanic인 랜덤 포레스트 모델을 사용한 예측

# 다운로드: https://raw.githubusercontent.com/logpresso/dataset/main/titanic/train.csv table titanic\_test | rforest model=rforest\_titanic | eval \_guess = if(\_guess=="0", "사망 ", "생존")

서브쿼리로 트레이닝 셋을 사용하는 경우

table titanic\_test | rforest target=Survived Pclass, Sex, Age, Fare, Embarked [ csvfile /test/train.csv | eval Age=double(Age), Fare=double(Fare), CanbinLetter=nvl(substr(Cabin, 0, 1), "--"), TicketType=if(isnull(long(Ticket)), substr (Ticket, 0, indexof(Ticket, " ")), "--") | rex field=Name ", (?<Title>[^.]+)" | eval Survived = if(Survived=="0", " 사망 ", "생존") ]