

# ランタイム環境・モデル投稿について

## 目次

- ランタイム環境概要
- モデル提出方法
- FAQ

# ランタイム機能を採用したコンペティションとは

課題に対する予測結果のファイルではなく、モデル一式のソースコードを投稿していただき、共通の計算環境（ランタイム環境）にて推論を行うコンペティションです。

	ランタイム環境非採用	ランタイム環境採用
モデルの学習	任意の環境	任意の環境
モデルの推論	任意の環境	ランタイム環境
提出物	予測結果ファイル (csvファイル等)	予測用コード一式

本コンペティションはこちら

コンペティションによってランタイム環境の仕様は異なります。

本コンペティションでは、以下の環境を設定しております。

- Docker : continuumio/anaconda3:2019.03(Customized)
- vCPU : 3
- メモリ : 16GB  
(ただし、実際の推論にお使いいただけるメモリは12GB程度です)
- 推論時間 : 最大3時間に制限
- 外部通信 : 外部インターネットへの通信は不可

# 提出ファイルが満たすべきディレクトリ構成

3

**model ディレクトリ**、**src ディレクトリ**、src ディレクトリ直下の **predictor.py** の3点が必須となっております。

これらを以下に示すように配置した上で、その他の必要な内容物をまとめて zip 化したファイルを提出してください。

├ model	必須 学習済モデルを置くディレクトリ
└ ...	
├ src	必須 Python のプログラムを置くディレクトリ
├ predictor.py	必須 最初のプログラムが呼び出すファイル
└ ...	その他のファイル (ディレクトリ作成可能)
└ requirements.txt	任意

図 提出ファイルのディレクトリ構成

推論を行うプログラムは predictor.py に記載をしてください。

predictor.py が満たすべき要件は以下の通りです。具体的な記載方法の詳細はデータページからダウンロードできる sample\_submit.zip ファイル、あるいはチュートリアルの内容をご確認ください。

## ScoringService クラス内に以下2つのメソッドを記載

### ■get\_model : モデルを取得するメソッド

- ✓クラスメソッドであること
- ✓引数 model\_path (str 型) を指定すること
- ✓正常終了時は返り値を True (bool 型) とすること

### ■predict : 推論を実行するメソッド

- ✓クラスメソッドであること
- ✓引数 input (str 型) を指定すること

## (参考) ランタイム環境上でモデルの学習を行いたい場合

get\_model が必ず True を返すようにして、predict メソッド内で特徴量生成やモデルの学習・保存が行われるように記述してください。

# predictor.py の処理実行順序とエラーメッセージ種別

5

predictor.py は get\_model メソッド、predict メソッドの順で実行されます。

get\_model



predict

投稿内容にエラーが発生した場合は、投稿済みファイル管理画面でもエラーメッセージが確認できますが、詳細なエラー内容についてはメールに記載されておりますので、そちらをご確認ください。

## エラーメッセージ種別

### エラーコード エラー発生原因

402 zip ファイルが解凍できなかった時

403 src または model ディレクトリが無かった時

411 requirements.txt にて異常があった時

412 推論処理がタイムアウトした時

416 モデルのロードができなかった時

417 プログラムのエラーが発生した時

ランタイム環境概要

モデル提出方法

FAQ

フォーラム等にお寄せいただいている質問の中から、件数の多さや、その内容からして共有すべきと思われる以下5点について、回答をご紹介します。

## Questions

1. get\_modelでNo module named 'sklearn.ensemble.\_forest'と出てしまいます。
2. Runtime計測に失敗しました。417: Killedとは具体的に何が原因でしょうか。
3. ランタイム環境にライブラリを追加する方法は？
4. 予測値に NaN が含まれていた場合はどのように処理されるのでしょうか？  
omitされるのか、平均値を埋め込むのか、あるいは NaN があった場合失格となるのでしょうか？
5. predict 関数の出力はどのような形式にすべきか

## Question

get\_modelでNo module named 'sklearn.ensemble.\_forest'と出てしまいます。

## Answer

Anaconda3-2019.03をお使いか、確認をお願いします。

ランタイム環境と同一の docker を利用せずにモデルを作成する場合は、利用するライブラリのバージョンをランタイム環境のものと合わせていただく必要があります。

特に scikit-learn のバージョンは、ランタイム環境では 0.20.3 であることにご注意ください。



## Question

Runtime計測に失敗しました。417: Killedとは具体的に何が原因でしょうか。

## Answer

詳細なエラー内容はメールの記載内容をご確認ください。

ただし、この場合はメモリエラーの可能性が高いです。推論にお使いいただけるのは約12 GB 程度であることにご留意ください。

また、例外処理のコードを記載いただいている場合は、そのコードを除外することで詳細なエラーメッセージが確認可能となっております。

## Question

ランタイム環境にライブラリを追加する方法は？

## Answer

requirements.txt に記載したライブラリは、ランタイム実行時に pip でインストールされます。

ただし、ライブラリを追加する際は、ランタイム環境の docker イメージ※にて、インストール及び使用可能であるかをご確認ください。

※ [continuumiio/anaconda3:2019.03\(Customized\)](#)

## Question

予測値に NaN が含まれていた場合はどのように処理されるのでしょうか？omit されるのか、平均値を埋め込むのか、あるいは NaN があった場合失格となるのでしょうか？

## Answer

NaN が含まれる場合、NaN を含む箇所も含めたエラーメッセージが返ります。詳細なエラー内容はメールの記載内容をご確認ください。

このため、プライベートリーダーボードの期間にてモデルの予測値に NaN が含まれていた場合は、順位評価対象外となります。

提出の際には予測値に NaN、および  $\pm\infty$  を含まないようにご注意ください。

## Question

predict 関数の出力はどのような形式にすべきか

## Answer

「index」、「最高値への変化率」、「最安値への変化率」の3カラムを出力してください。また、ヘッダーは付けないでください。

「index」はデータ基準日と銘柄を-（ハイフン）で繋いだ形式とします。  
例えば、2016年5月9日発表の銘柄1301の場合、index は **2016-05-09-1301** としてください。

出力するコードについては、チュートリアルをご参考ください。

ご清聴ありがとうございました