

NEC Advanced Analytics Platform V1.5 マネージドサービス サイジングガイド (簡易版)  
異種混合学習編 (第1版)

NEC Advanced Analytics Platform V1.5 マネージドサービスで異種混合学習(SAMPO/FAB)もしくは異種混合学習(sklearn-fab)を使用して学習・分析を行うにあたって、入力データのデータ量(「サンプル数×属性数」の量)に応じた契約プランの目安を表1に示します。

表1 入力データ量と契約プランの目安

入力データ量 (サンプル数×属性数 の量)	選択するプラン(目安)
40万まで	エントリー エントリープラス(1コア)
80万まで	エントリープラス(4コア)
上記より多い	スタンダード

前提条件：

- ・ 使用している AACluster の CPU コア数と同数の並列度を指定して学習処理を行います。
- ・ 全モデル作成数を 10、ランダムリスタート数を 10、学習結果保存世代数を 5 と仮定しています。これらの数値を仮定より大きくする場合は、必要なストレージが増えますので、扱う入力データ量を減らす、あるいは上位のプランの契約やストレージ容量追加オプションの検討が必要です。

注意事項：

- ・ 分析で使用する入力データ以外のファイルがストレージにある場合、扱える「サンプル数×属性数」が減ります。
- ・ 二値展開を行う属性数が多い場合、およびその属性の中の値の種類数が多い場合、メモリー使用量が極端に大きくなる場合があります。
- ・ データの独自の加工を行う場合、その加工において必要なファイル容量は上記目安値に含めていないため、扱える「サンプル数×属性数」が減ります。
- ・ 本書で示す数値はあくまで目安ですので、お客様の分析データや分析方法によってはメモリー使用量やストレージ使用量が想定を上回る可能性があります。エントリープラン、エントリープラスプランでは後からメモリーを追加できないため、およびエントリープランでは後からストレージを追加できないため、不安な場合は1ランク上のプランを選択される等ご検討ください。
- ・ 本書で示す各数値はあくまで目安ですので、扱うデータ内容や分析方法によっては扱えるデータ量が大きく異なる場合があります。より確からしい見積もりを出すためには、本番運用と同等の特性を持つサンプルデータを用いるなどして実測することを推奨します。

用語説明：

サンプル数	学習処理の入力データの行数。
属性数	学習処理の入力データの列数。例えば、年齢、性別などの属性の数。
並列度	同時に処理するワーカースレッド数。実行時に「--worker」オプションで指定します。
AACluster	本マネージドサービスが提供する仮想分析環境(コンテナ)
ランダムリスタート数	異種混合学習エンジンが持つ初期解のランダム性を考慮して試行する学習の回数の設定値
二値展開	カテゴリ変数(性別、職業など一般に数や量で測れない変数)を値として持つ属性の各属性値を0または1の二値に変換すること。

本書について：

- ・ より詳細な見積もりが必要な場合は、「NEC Advanced Analytics Platform V1.5 マネージドサービス サイジングガイド」を参照してください。
- ・ 本書を事業者の許可なくコピーおよびその配布、ホームページへの掲載を禁じます。