

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)
RAPID 機械学習 マッチング版編 (第1版)

本シートでは RAPID 機械学習のマッチング版の分析に必要なメモリー、ストレージ容量の参考値を説明します。

マッチング版には「マッチング機能」と「フィルタリング機能」があります。いずれの機能も分析するテキストをテキスト列、数値列、データ行に分けて分析します。

表 1 にはマッチング機能における 6 ケースの分析データ・パターンにおけるメモリー、ストレージ使用量の実測値を示します。表 2 には同様にフィルタリング機能における実測値を示します。

これらの情報をもとにお客様が分析されようと考えている分析データ・パターンのデータ条件と照らし合わせて必要なメモリー、ストレージ量を推測して見積もり、最適なプランを選択します。

表 1 学習実行時の実測値 (マッチング機能)

| No | データ条件 | | | | 使用リソース | | |
|----|---------|--------|------|---------|----------|---------|----------|
| | テキスト文字数 | テキスト列数 | 数値列数 | データ行数 | 入力データサイズ | メモリー使用量 | ストレージ使用量 |
| 1 | 1,000 | 0 | 50 | 10,000 | 3.9MB | 94MB | 116KB |
| 2 | 1,000 | 1 | 50 | 10,000 | 64MB | 498MB | 176KB |
| 3 | 1,000 | 10 | 50 | 10,000 | 601MB | 4,201MB | 724KB |
| 4 | 1,000 | 1 | 10 | 10,000 | 61MB | 488MB | 156KB |
| 5 | 1,000 | 1 | 100 | 10,000 | 68MB | 513MB | 200KB |
| 6 | 1,000 | 1 | 50 | 100,000 | 635MB | 1,566MB | 176KB |

表 2 学習実行時の実測値 (フィルタリング機能)

| No | データ条件 | | | | 使用リソース | | |
|----|---------|--------|------|--------|----------|---------|----------|
| | テキスト文字数 | テキスト列数 | 数値列数 | データ行数 | 入力データサイズ | メモリー使用量 | ストレージ使用量 |
| 1 | 1,000 | 0 | 50 | 10,000 | 2.4MB | 84MB | 12KB |
| 2 | 1,000 | 1 | 50 | 10,000 | 33MB | 148MB | 44KB |
| 3 | 1,000 | 10 | 50 | 10,000 | 304MB | 702MB | 300KB |
| 4 | 1,000 | 1 | 10 | 10,000 | 31MB | 144MB | 40KB |
| 5 | 1,000 | 1 | 100 | 10,000 | 35MB | 151MB | 48KB |
| 6 | 1,000 | 1 | 50 | 10,000 | 325MB | 734MB | 44KB |
| 7 | 1,000 | 5 | 50 | 10,000 | 1,566MB | 2,774MB | 160KB |

注意事項：

- 表 1、表 2 は、学習モードが分類の場合のメモリー使用量、ストレージ使用量の実測値ですが、回帰の場合もほぼ同様のメモリー使用量、ストレージ使用量となります
- ストレージ容量としては、表 1、表 2 のストレージ使用量に入力データ (※1) を加えてストレージ容量を見積もってください。さらに学習・予測処理を行う前に前処理としてデータの加工を行う場合、入力データサイズをさらに加えて見積もってください (※2)。
 ※1：表 1、表 2 ストレージ使用量には入力データが使用するストレージ容量を含みません。
 ※2：前処理の実行で入力データサイズとほぼ同等のストレージを消費します。
- テキスト列数が増加すると、メモリー使用量が大きく増加することに注意してください。
- 本書で示す実測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。エントリー、エントリープランでは後からメモリーを追加できない

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。

用語説明：

| | |
|-----------|---|
| マッチング機能 | データ A、データ B の 2 つ組を入力データとして、両者の適合度合いをマッチングスコアとして数値化する機械学習アルゴリズム全般を指します。RAPID 機械学習 マッチング版の一つの機能です。 |
| フィルタリング機能 | データの 1 つ組を入力データとして、その適合度合いをフィルタリングスコアとして数値化する機械学習アルゴリズム全般を指します。RAPID 機械学習 マッチング版の一つの機能です。 |

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。