

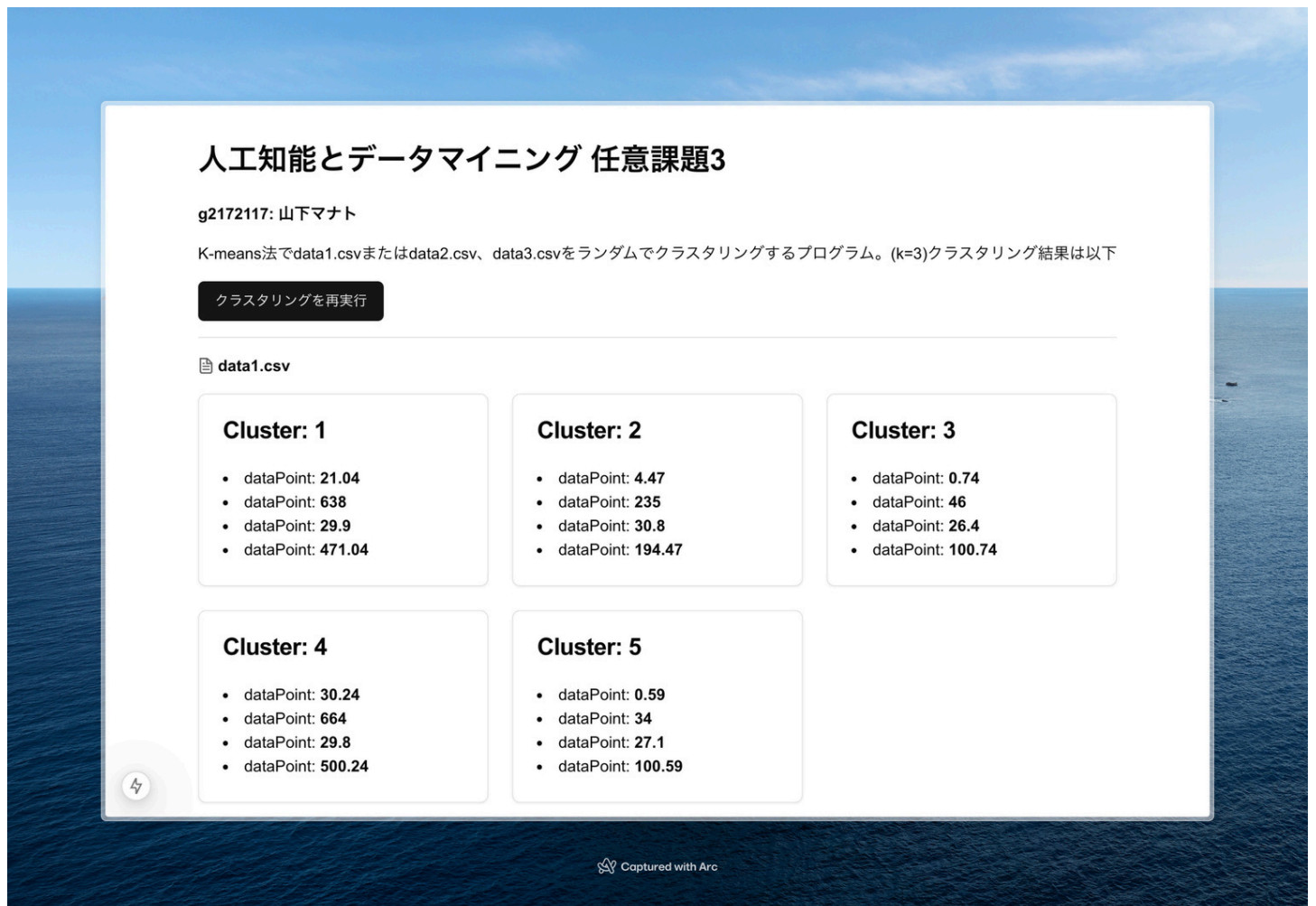
太陽光発電と消費電力に関するレポート

情報システム学科 4年生
g2172117: 山下マナト

<https://kmeans-clustering.vercel.app>

<https://github.com/ManatoYamashita/kmeans-clustering>

1. クラスタリング結果



クラスタリングの手法

本レポートでは、Next.jsを用いてWebAppを作成して、K-means法 ($k=3$) を用いてクラスタリングを実施しました。K-means法は、指定したクラスタ数 (k) に基づき、データを k 個のクラスタに分割する手法です。各クラスタの中心点（重心）からの距離が最小となるようにデータをグループ化します。K-means法の特徴として、各クラスタの境界が線形で決まるため、類似度の高いデータが自動的に近いクラスタに割り当てられる点が挙げられます。

クラスタ数 (k) の選択

クラスタ数 k は3に設定し、データを3つのグループに分けました。この k の設定は、太陽光発電量、日射量、気温、消費電力量に基づき、データに含まれる類似性に応じて適切に分けられるように選択しました。

使用したデータ

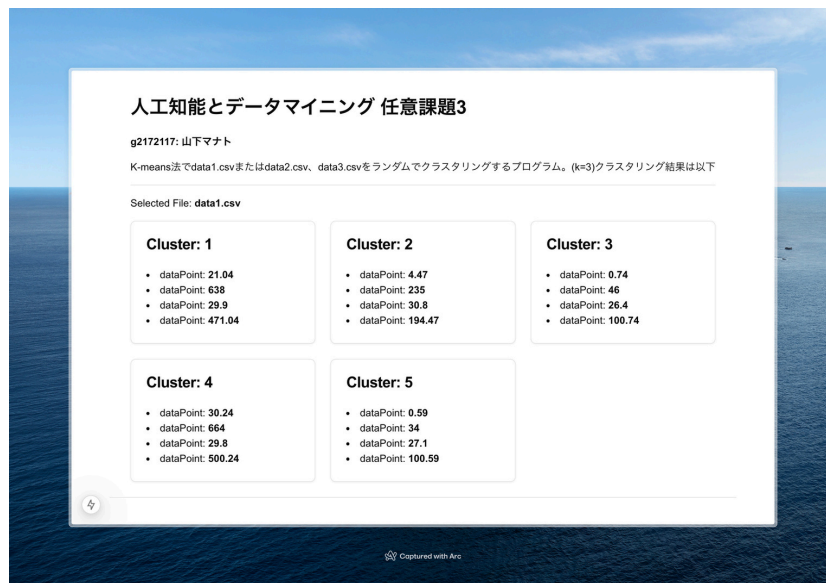
以下の3種類のデータセットを使用して、クラスタリングを実行しました。

- a. **data1**: 生データ。各測定項目（発電量、日射量、気温、消費電力量）が含まれており、値はそのままの単位で記録されています。
- b. **data2**: 正規化データ（上限値が1）。各測定項目が上限値1になるようにスケーリングされています。このデータは測定値の比較や比率に着目する際に有用です。
- c. **data3**: 正規化データ（下限0, 上限1）。各測定項目が0から1の範囲に収まるようにスケーリングされており、データ範囲の差異を最小化するために利用します。

クラスタリング結果

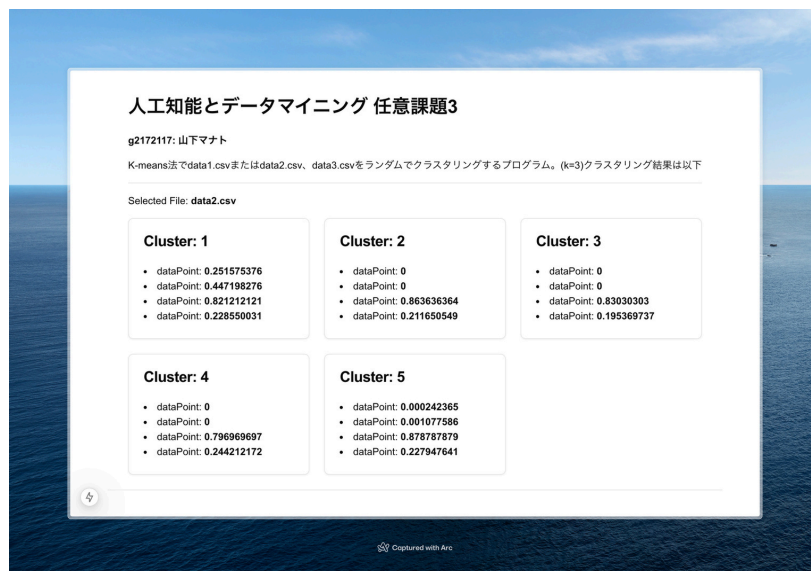
各データセットのクラスタリング結果は以下の通りです。

- data1 のクラスタリング結果



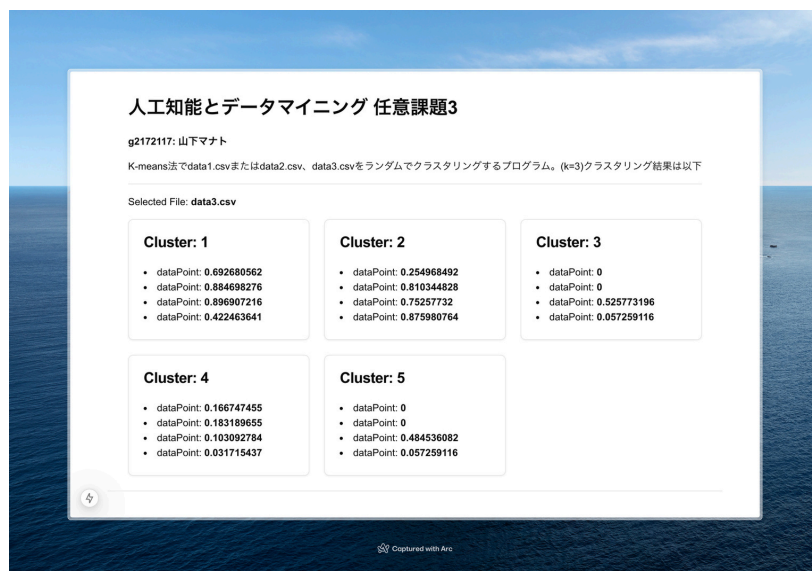
```
[ [ 0, 0, 27.4, 120 ],
  [ 0.39, 26, 27.3, 120.39 ],
  [ 2.38, 74, 27.4, 122.38 ],
  [ 6.2, 134, 27.5, 206.2 ],
  [ 7.15, 188, 25.9, 377.15 ]
]
```

○ data2 のクラスタリング結果



```
[ [ 0, 0, 0.83030303, 0.195369737 ],
  [ 0.009452254, 0.028017241, 0.827272727, 0.196004689 ],
  [ 0.057682986, 0.079741379, 0.83030303, 0.19924457 ],
  [ 0.150266602, 0.144396552, 0.833333333, 0.335710332 ],
  [ 0.173291323, 0.202586207, 0.784848485, 0.614030803 ]
]
```

○ data3 のクラスタリング結果



```
[ [ 0, 0, 0.422680412, 0.037789849 ],  
  [ 0.009452254, 0.028017241, 0.412371134, 0.03854915 ],  
  [ 0.057682986, 0.079741379, 0.422680412, 0.042423534 ],  
  [ 0.150266602, 0.144396552, 0.432989691, 0.205614937 ],  
  [ 0.173291323, 0.202586207, 0.268041237, 0.538442069 ]  
]
```

2. 考察

クラスタリングの結果を通じて、データの特徴に基づいたパターンが浮き彫りになりました。

データ特性とクラスタリング

data1（生データ）では、データの範囲がそのまま反映され、特に発電量や消費電力量の影響がクラスタリングに大きく現れています。一方、data2やdata3の正規化データでは、測定項目間の比較がしやすくなり、特定の項目に依存することなく、全体の傾向を反映するクラスターが形成されています。data2とdata3の違いとして、下限と上限を0~1に統一したdata3は、特に発電量の変化が大きく影響するクラスタリング結果となり、各測定項目が均等に考慮される形となっています。

K-means法の適用に関する評価

K-means法の特徴上、データの分布がクラスター境界に影響を与えます。特に発電量や消費電力量の値が大きく異なる場合、クラスターが分散するため、数値のスケーリングが重要となります。今回のように、元データと正規化データの両方でクラスタリングを行うことで、項目ごとの影響を比較しやすくなり、特徴の理解が深まりました。

今後の改善点

- **クラスタ数の最適化:** $k=3$ でクラスタリングを行いました、異なるクラスタ数を試すことで、さらにデータの傾向や特性を詳細に分析できる可能性があります。
- **他のクラスタリング手法の検証:** 今回はK-means法を使用しましたが、K-medoidこれに法により、クラスタリングの精度がさらに向上するかもしれません。

3. プログラムソース

プログラムは以下のGithubリポジトリに格納しております。

- Github
<https://github.com/ManatoYamashita/kmeans-clustering>
- WebApp
<https://kmeans-clustering.vercel.app>

使用技術

- Next.js - AppRouter(v15)
- TypeScript - Javascript superset
- PapaParse - CSVファイルのparseライブラリ
- kmeans-js - K-meansクラスタリングjsライブラリ

以上です

人工知能とデータマイニング 任意課題3

g2172117@tcu.ac.jp

東京都市大学 メディア情報学部 情報システム学科

山下マナト

<https://manapuraza.com>