

背景と目的

- ホタルイカの「身投げ」は、春（2～5月）に富山湾の海岸へ大量に押し寄せる産卵行動。
- 幻想的な光景を生み、地域の文化として親しまれている。
- 現在の情報源は、個人の経験則や掲示板サイトが中心。
- データに基づく高精度な予測システムはまだ存在しない。

本研究の目的は、具体的には、以下の2点を達成することを目的とする。

- 過去の身投げ実績と気象・潮汐データを用いて、機械学習で身投げ量を予測するモデルを構築する。
- 予測モデルを組み込んだAPIや掲示板機能を備えたWebアプリを開発し、ホタルイカ掬いの利便性を高める。

研究方法

本研究は、「1, データセット構築」「2, 機械学習モデル訓練」「3, Webサイトの開発」の3フェーズで実施した。

1, データセット構築

過去10年間の2～5月（約1200日分）について、以下のデータを収集

- ホタルイカの身投げ量
- 気象データ
- 月齢データ



身投げ量データ

掲示板の口コミをAIで6段階に分類



その値の平均値をその日の身投げ量とする

1, 使用データと前処理

使用データ

目的変数: ホタルイカ身投げ量
説明変数: 月齢、気温、降水量、風速、風向、日付データ

周期性データの変換

課題: 周期データは数値の大小が距離を反映しない。
解決策: sin/cos変換で周期性を表す。

月齢 (周期 $T = 29.53$):

$$\text{moon_age_sin} = \sin\left(\frac{2\pi \cdot \text{moon_age}}{29.53}\right)$$

$$\text{moon_age_cos} = \cos\left(\frac{2\pi \cdot \text{moon_age}}{29.53}\right)$$

風向データの平均化 (ベクトル合成)

課題: 風向は循環データのため単純平均できない。
解決策: sin/cosでベクトル平均し角度を求める。

風向データ d_i (0～15) を角度 θ_i に変換しその角度のサイン (y成分) とコサイン (x成分) の平均 (\bar{S}, \bar{C}) を求める

$$\bar{S} = \frac{1}{N} \sum_{i=1}^N \sin(\theta_i) \quad \bar{C} = \frac{1}{N} \sum_{i=1}^N \cos(\theta_i)$$

平均角度 $\bar{\theta}$ を復元し、0～15のインデックスに戻す

$$\bar{\theta} = \text{atan2}(\bar{S}, \bar{C})$$

2, 機械学習モデルの訓練

2, モデル構築

使用モデル: LightGBM

勾配ブースティング決定木を用いた高速・高精度な機械学習モデル

勾配ブースティングとは？

モデルの「間違い」を修正する決定木 f_t を逐次的に追加していく手法

1, モデル更新

t 回目の予測 F_{t-1} は、前の予測 F_{t-1} に新しい木 f_t を足して作られる

$$F_t(x) = F_{t-1}(x) + f_t(x)$$

2, 間違いの計算

現在の予測 F_{t-1} がどれだけ間違っているかを、損失関数 L の微分で計算

$$g_i = \left[\frac{\partial L(y_i, F)}{\partial F} \right]_{F=F_{t-1}(x_i)}$$

3, 新しい木の決定

新しい木 f_t は、この「間違い」(の逆方向 $-g_i$) を予測するように学習

$$f_t(x) \approx -g_i$$

これを繰り返し、モデルは徐々に間違いを減らしていく。

3, モデルの評価

予測精度を客観的に評価するため、以下の指標を用いた。

決定係数

分散をどの程度説明できているかを示す指標

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

平均絶対誤差

予測誤差の平均的な大きさ

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

二乗平均平方根誤差

予測誤差の大きさを示す代表的な指標

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

n : データ数, y_i : 実測値, \hat{y}_i : 予測値, \bar{y} : 実測値の平均

3, Webサイトの開発

サイトURL

<https://bakuwaki.jp/>

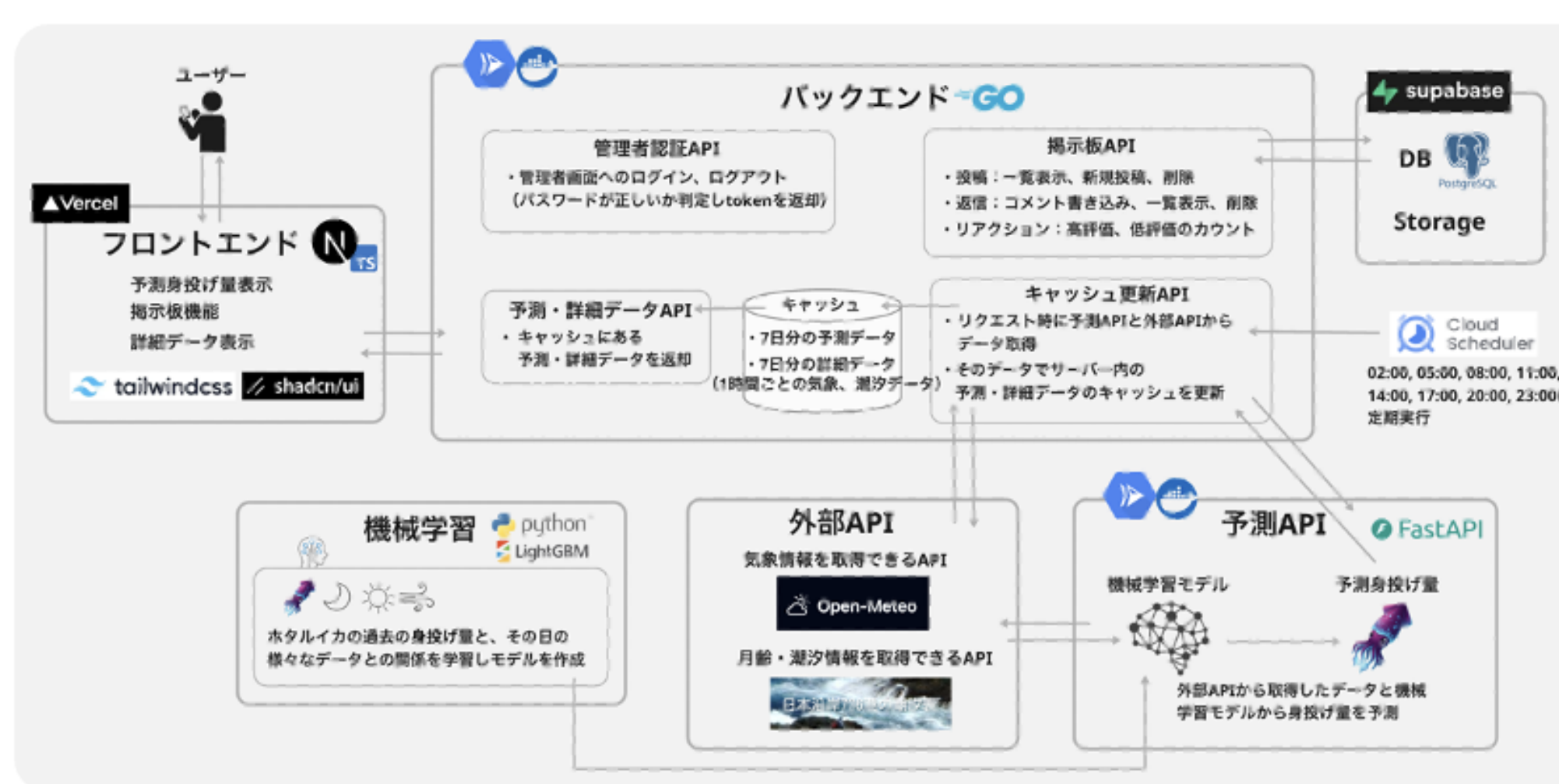
シーズン中の表示: <https://bakuwaki.jp/preview/>

機能

- 身投げ予測: 今日から1週間後までのホタルイカの発生量を予測
- 詳細データ: 気温、風速、潮位などの関連データが確認できる
- 掲示板: ホタルイカの情報を共有できる

使用技術

- フロントエンド: Next.js, TypeScript
- UI: Tailwind CSS, shadcn/ui
- バックエンド: Go, FastAPI
- DB・ストレージ: Supabase (PostgreSQL)
- インフラ: Docker, Vercel, GCP(Cloud Run)



結果と考察

予測モデルの評価

構築したLightGBMモデルの精度を評価するため、全データセットの最後の20%をテストデータとして使用し、予測値と実測値を比較した。

評価指標:

決定係数 (R^2): 0.7008

平均絶対誤差 (MAE): 0.0721

二乗平均平方根誤差 (RMSE): 0.1013

予測結果の可視化: 予測値 (赤色) と実測値 (青色) を時系列で比較したグラフ(記事の「予測結果と実測値の比較」図を参照)を作成した。

特徴量の重要度: 予測への寄与度が高い特徴量として、temperature_std(気温の標準偏差)、precipitation_sum(降水量)、day_of_year_sin(日付の周期性)、moon_age_cos(月齢の周期性)、wind_speed_std_lag1(1日前の風速の標準偏差)などが上位となった。

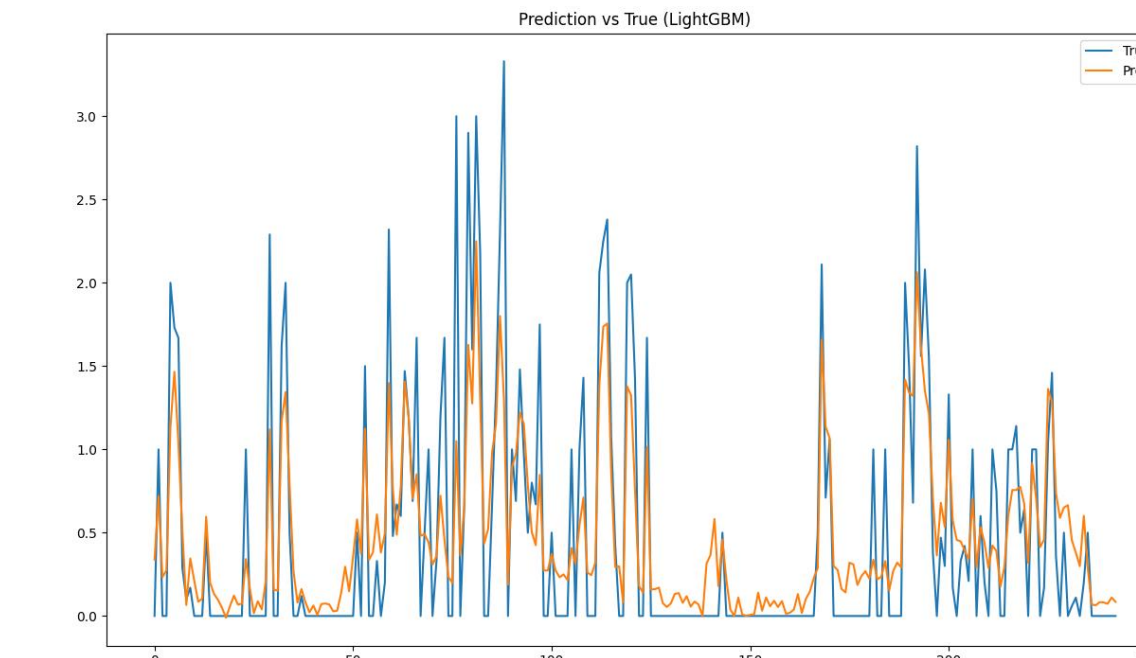
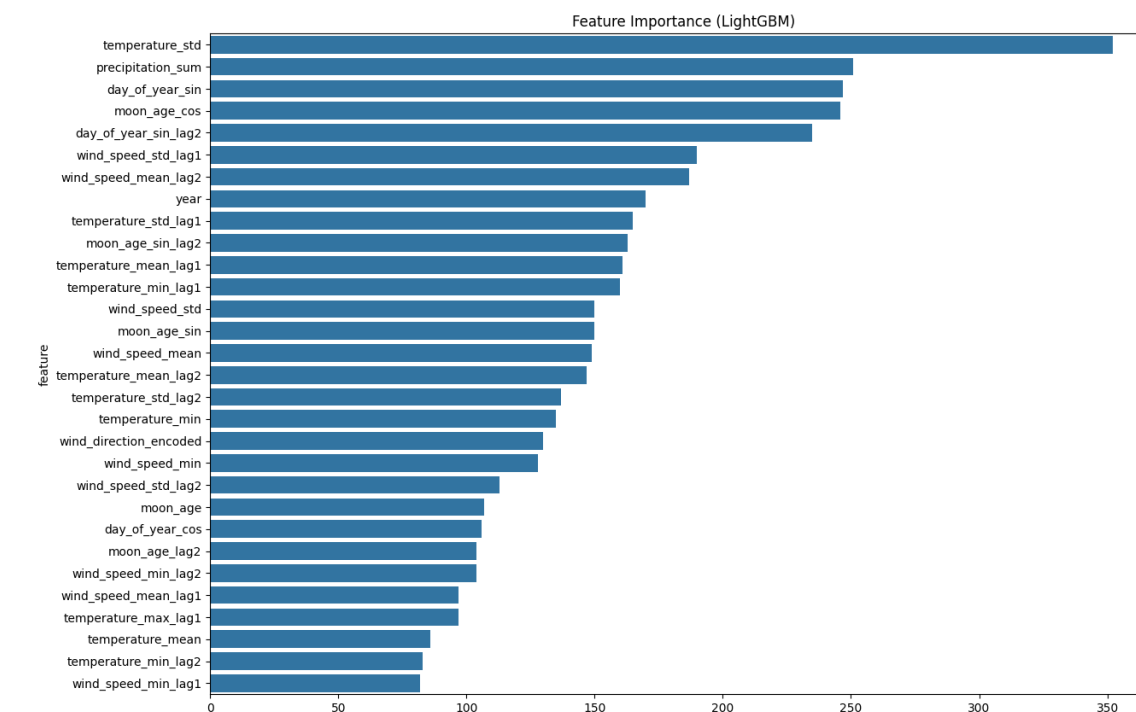
考察と課題

構築したLightGBMモデルのテストデータに対する決定係数(R^2)は **0.7008** となり、一定の予測精度を達成した。特徴量の重要度分析からは、季節性(day_of_year)、月齢、気温、降水量といった、従来経験的に重要とされてきた要因が予測に寄与していることを定量的に確認できた。

一方で、モデルには以下の課題も明らかになった。

予測の精度: 全体の増減傾向は捉えられているものの、突発的な「爆湧き」のようなピーク値の予測精度には改善の余地がある。

外部データへの依存: モデルの予測精度は、入力データとして利用する外部の気象予報API自体の精度に直接依存する。気象予報が不正確な場合、本モデルの予測精度も低下する。



今後の予定

- 予測精度の向上: 海岸ごとの位置情報を特徴量として加え、より空間解像度の高い局所的な予測モデルを構築する。また、より高精度な気象・海象データを導入する。
- Webサイトの機能拡充: オフシーズンでもユーザーが訪れる価値のあるコンテンツ(例: 過去の身投げデータの統計・可視化ページなど)を追加し、サイトの継続的な利用を促進する。

⑧参考文献（任意）

⑧参考文献（任意） の内容を記載してください。