

GeoSciAI 2025 コンペ テクニカルレポート

Taka.N

2025 年 5 月 12 日

1 概要

本レポートでは、2009～2019 年の太陽風・地磁気観測データを用いて 24h 先の Dst 指数を回帰する機械学習モデルを構築した結果をまとめる。物理ドメイン知識に基づく磁気圏-太陽風結合関数とフーリエ級数特徴を LightGBM へ投入し、公開テスト期間（2024/5/2 ～ 5/31）で **49.169 nT** の RMSE を達成した。

学習に使用したコードは https://github.com/Taka0007/GEOSciAI2025_sub で公開している。（提出期限終了後に Public レポジトリにする予定。）

2 データの使用方法

2.1 データセット

- 訓練データ: geosci.ai2025_sw_train.csv
- テストデータ: geosci.ai2025_sw_all_test_timeshift.csv

2.2 Training/Validation 分割

時系列リークを避けるため、**Walk - Forward 型 TimeSeriesSplit (n=5)** を採用した。各 fold でおおむね 1 年分（ ≈ 8760 サンプル）を検証に用いている。

3 特徴量エンジニアリング

3.1 特徴量の増強

与えられた特徴量の演算を行い、11 種の指標を算出した。

特徴量の算出に使用したコードを下記 1 に示す。

Listing 1 特徴量算出スクリプト

```
1 def add_domain_features(df):  
2     df["Bs"] = df["bz (nT)"].clip(upper=0).abs()  
3     df["BT"] = np.sqrt(df["by (nT)"]**2 + df["bz (nT)"]**2)
```

```

4     df["theta_c"] = np.degrees(np.arctan2(df["BT"], df["bx (nT)"]))  $\mu$ 
5     0 = 4*np.pi*1e-7
6     df["epsilon"] = 1e3 * (df["vsw (km/s)"] * (df["BT"]**2) * (np.sin(np.
        radians(df["theta_c"])/2)**4)) /  $\mu$ 
7     df["phi_newell"] = (df["vsw (km/s)"]**(4/3)) * (df["BT"]**(2/3)) * (np
        .sin(np.radians(df["theta_c"])/2)**(8/3))
8     df["Ey_calc"] = -1e-3 * df["vsw (km/s)"] * df["bz (nT)"]
9     df["Q_burton"] = 0.20 * df["Ey_calc"] - 1.5
10    df["Dst_star"] = df["dst (nT)"] - 7.26 * np.sqrt(df["pdyn (nPa)"]) + 11
11    return df

```

3.2 時系列派生特徴

- ラグ特徴: 1-24 h のタイムラグ
- ローリング統計: 3/6/12 h の平均・標準偏差・最小・最大
- フーリエ級数: 周期 (24,12,6,648,8766) h の sin/cos (次数 $k = 1$)

最終的な総特徴量数は、278 次元となった。

4 モデル構造と学習方法

4.1 LightGBM 回帰

Gradient Boosting Decision Tree 実装である LightGBM4.3 を採用。目的関数は RMSE, ハイパーパラメータは Optuna (40 trials, TPE サーチ) で最適化した。主な最適値は以下:

パラメータ	値
num_leaves	779
max_depth	15
learning_rate	0.011
feature_fraction	0.98

Storm イベント ($\text{Dst} < -50 \text{ nT}$) を 3 倍に重み付けし、外れ値寄与を低減。

5 結果

5.1 交差検証

5-fold 平均 RMSE は **13.19 nT**。

5.2 テスト期間 (2024-05-02-31)

- RMSE: 49.169 nT

- 図1 に真値と予測値の時系列を示す。

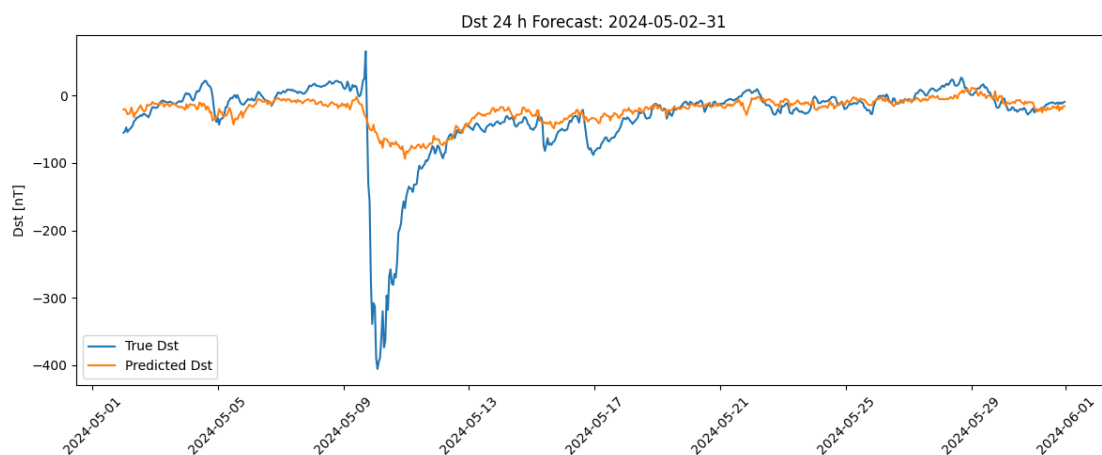


図1 2024年5月2日～31日における Dst 真値（青）と予測（橙）

6 結論と考察

物理指標＋フーリエ周期特徴を組み合わせた LightGBM により、公開データのみで 49.169 nT のテスト RMSE を達成した。フーリエ特徴を加えることで若干の RMSE 改善が見られたが、Dst の急落には対処できなかった。今後はドメイン知識を生かした特徴量を加えるなどの方面で工夫を凝らしていきたい。