The 7th International Conference on Deep Learning in Computational Physics,
June 21-23, 2023

Classification Approach to Prediction of Geomagnetic Disturbances

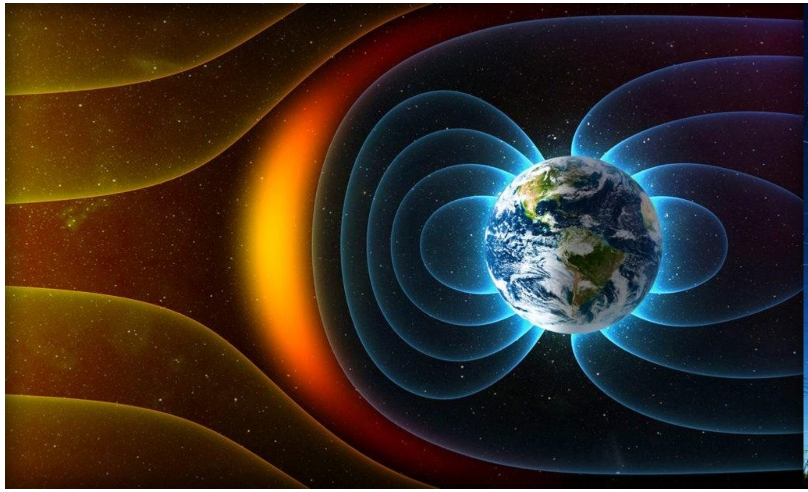
Gadzhiev I.M., Barinov O.G., Myagkova I.N., Dolenko S.A.

SINP MSU, Moscow

Исследование выполнено за счёт гранта Российского научного фонда, проект № 23-21-00237.

Prediction of Geomagnetic Disturbances

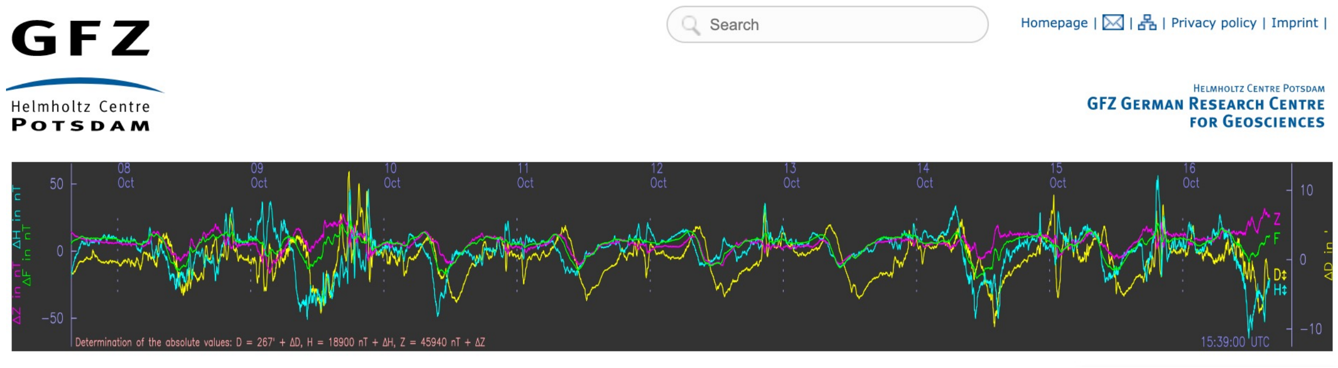
- Geomagnetic disturbances are one of the most important factors in space weather.
- They can cause disruption to radio communications, pipelines, power lines and electrical networks.
- Disturbance prediction can help to handle these problems.



Planetary index Kp

One of the most used geomagnetic indices is the planetary index Kp ($Kp \in [0 \ldots 9]$, with a step of 1/3)

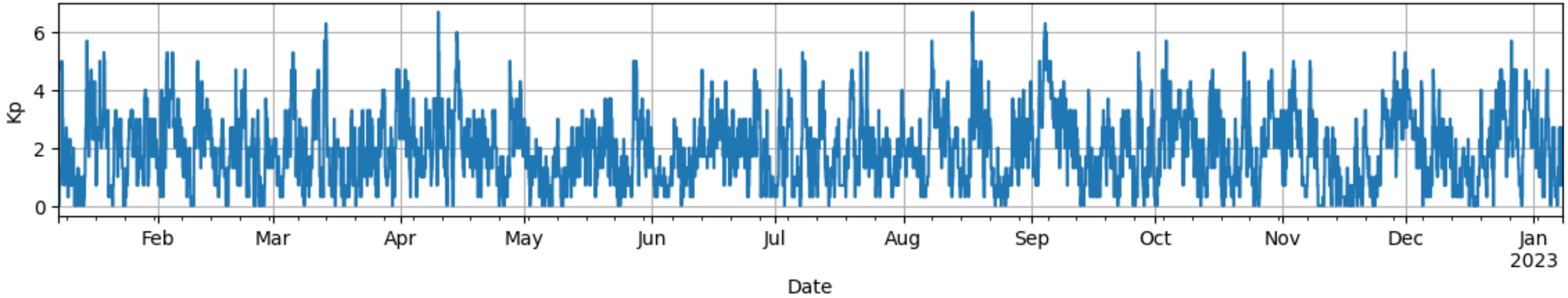
Kp is the weighted average of K-indices obtained at 13 observatories.
K-index is derived from the maximum fluctuations
of the horizontal components of the Earth's magnetic field,
observed on the magnetometer for 3 hours



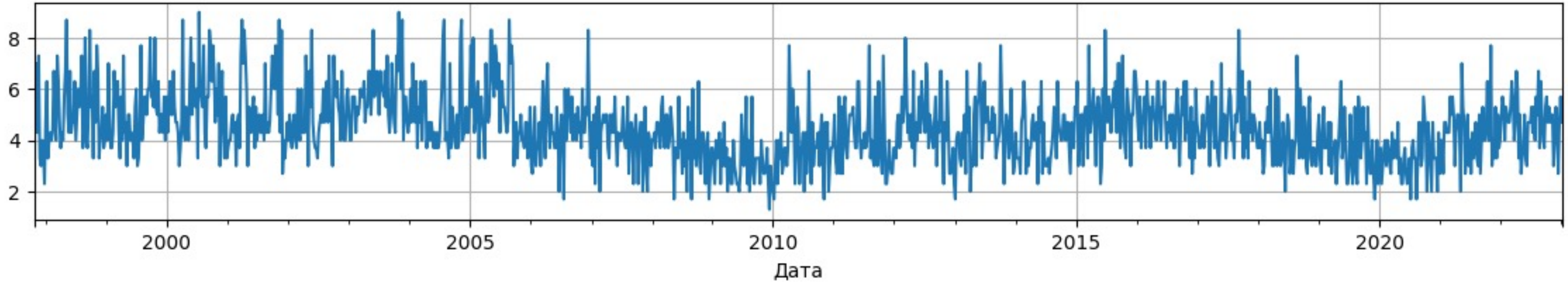
https://isdc.gfz-potsdam.de/kp-index/

Kp index dynamics

Dynamics for 2022 year



Weekly maximums since 1997



https://isdc.gfz-potsdam.de/kp-index/

Data for Kp forecasting

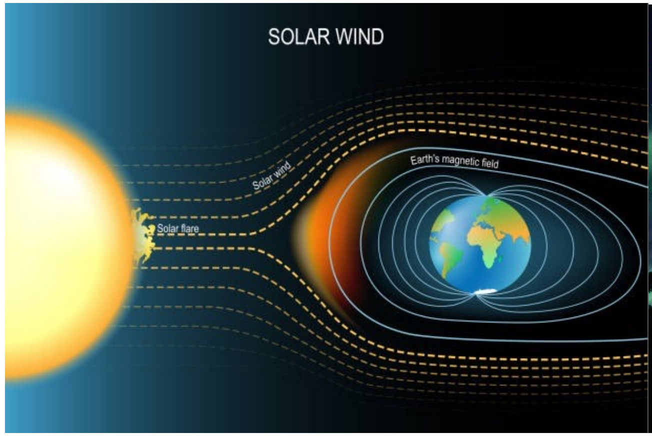
Feature	Description
Kp (previous)	Kp Index 3-hour
Dst	Dst Index
B_x	x component of the magnetic field
B_gsm_y	x component of the magnetic field (GSM)
B_gsm_z	y component of the magnetic field (GSM)
B_magn	Magnetic field module
SW_spd	Solar wind speed at the Lagrange point L1
H_den_SWP	Solar wind density at the Lagrange point L1
daySin	sin(2*π*[Day of a year]/365)
hourSin	sin(2*π*[Hour of a day]/24)
dayCos	cos(2*π*[Day of a year]/365)
hourCos	cos(2*π*[Hour of a day]/24)

Categories of Kp-index

The purpose is to classify the geomagnetic index category Kp.

1. Kp <=1.7 - undisturbed and weakly disturbed magnetosphere **(58%)**
2. 1.7 < Kp <= 3.3 - weak geomagnetic disturbances **(31%)**
3. Kp > 3.3 - medium and strong geomagnetic disturbances **(11%)**

Classes are imbalanced!



The target

Given the previous data, the target is to predict
the category of geomagnetic disturbance from 3 to 24 hours ahead
with step size 3 hours

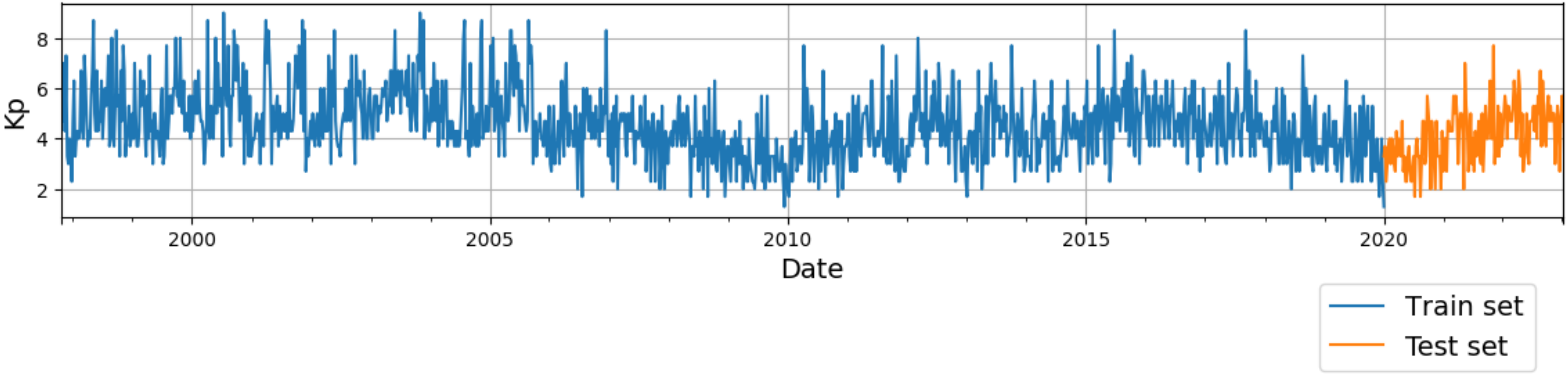
We test **a set of data preprocessing methods and machine learning models against this task**

For each horizon we select the best model. Previous research shows, that different models could succeed at each of the horizons!

Quality assessment

The range of values of all variables for 2020-2022 was used as a test set for assessing the quality of the models.

Smoothed dynamics of Kp divided into training and test sets:



The quality metrics (previous studies)

We previously used **F_1 macro-averaged** over all classes
to assess the quality of models and rank them (for each horizon)[1]:

$$< F_1 >= \frac{1}{3}\sum_{i=1}^{3} F_{1i}$$

$$F_{1i} = \frac{2 * Precision_i \ * Recall_i}{Precision_i + Recall_i}$$

All the models were compared to the **trivial “inertial” forecast**
– predicted category is equal to the last known one

[1] Use of Classification Algorithms to Predict the Grade of Geomagnetic Disturbance. Gadzhiev Ismail, Myagkova Irina, Dolenko Sergey. Advances in Neural Computation, Machine Learning, and Cognitive Research VI. Studies in Computational Intelligence. V 1064. P 426-435. Springer International Publishing AG. 2023.

The quality metrics (previous studies)

We previously used **F_1 macro-averaged** over all classes
to assess the quality of models and rank them (for each horizon):

$$< F_1 >= \frac{1}{3}\sum_{i=1}^{3} F_{1i}$$

$$F_{1i} = \frac{2 * Precision_i \ * Recall_i}{Precision_i + Recall_i}$$

All the models were compared to the **trivial “inertial” forecast**
– predicted category is equal to the last known one

The problem – this metric has equal weights for precision and recall!

Sometimes a model was chosen with recall < (recall of the trivial model)

The quality metrics (current setup)

Premises:

- Recall is more important than precision
- Classes 2 and 3 (disturbances) are more important than class 0 (no disturbances or weak disturbances)

Solution – **F_2 macro-averaged over 2 and 3 classes**

$$< F_2 >= \frac{1}{2} \sum_{i=2}^{3} F_{2i}$$

$$F_{2i} = \frac{5 * Precision_i \ * Recall_i}{4 * \ Precision_i + Recall_i}$$

All the models are compared to the **trivial "inertial" forecast** – predicted category is equal to the last known one

Data preprocessing

We tested several variants of data preprocessing

- Only current value of all features (no lags)
- 24h of lags with 3h step (8 lags)
- 24h of lags with 3h step for Kp and 1h step for other variables (8/24 lags)
- Different statitistics aggregated over rolling windows (agg):
  - Mean, Median, Quantiles, Max, Min of time series
  - Mean, Median, Quantiles, Max, Min of first difference of time series
  - Mean number of sign changes fo first difference if time series
  - Exponentially weighted averages

Models

- Linear model: **logistic regression (LR)**
- Gradient boosing: **LightGBM implementation (LGBM)**
- **Random Forest (RF)**
- **Perceptron**
- **Trivial model (TM)**
- Recurrent neural network: **LSTM** (with lags only)

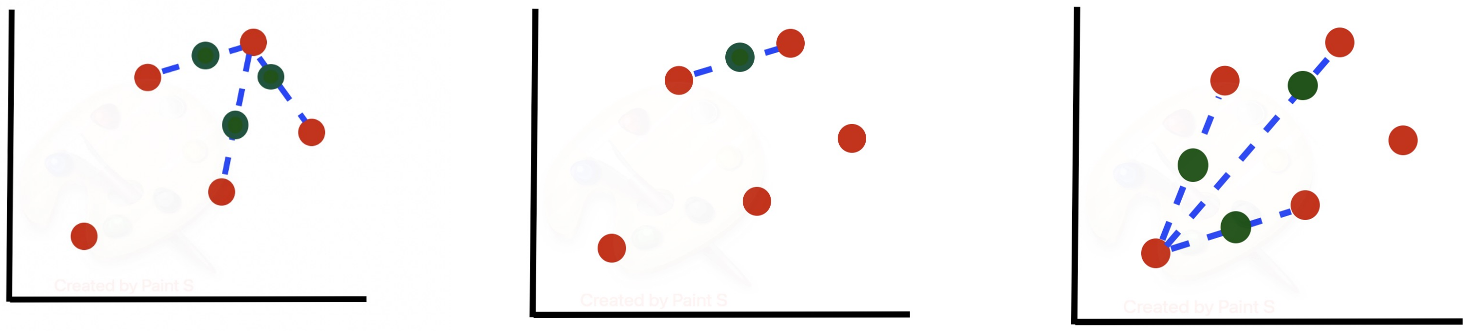
SMOTE

To handle class imbalance we use **SMOTE**

SMOTE - Synthetic Minority Over-sampling Technique.

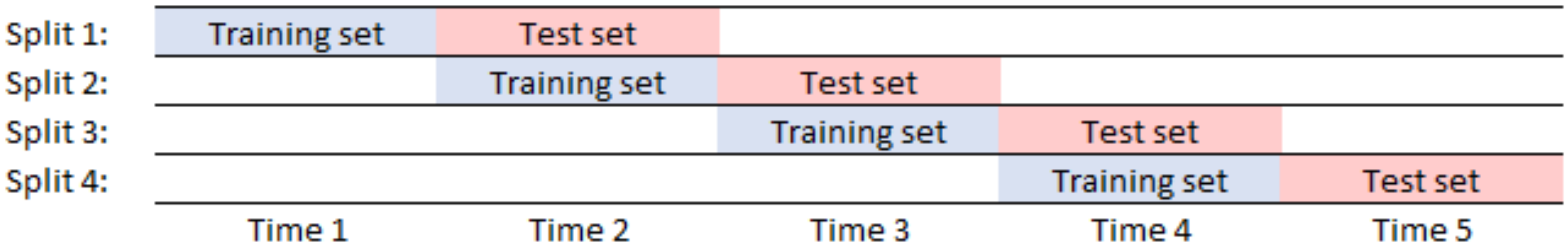
The essence of the technique is to level the balance by generating artificial examples in minority classes.

All listed models (except perceptron, LSTM)
are tested with and without SMOTE.



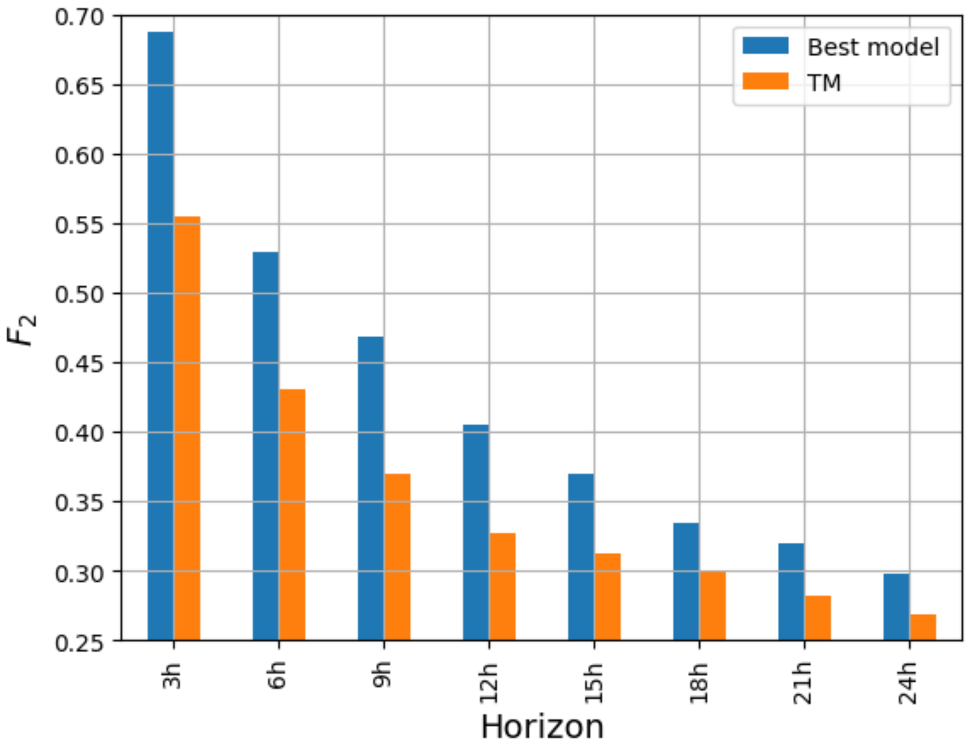
Cross-validation

For all models (in configurations with and without SMOTE), the cross-validation technique for time series (TSCV) was also tested to select the optimal hyperparameters.



Best models (all horizons)

Best score for each horizon
(compared to trivial model)



Best model for each horizon

3h - 8/24 lags SMOTE LR
6h - 8 lags SMOTE TSCV LR
9h - 8/24 lags SMOTE TSCV LR
12h - 8/24 lags SMOTE TSCV LR
15h - 8 lags SMOTE TSCV LR
18h - 8 lags SMOTE LR
21h - 8/24 lags SMOTE LR
24h - Agg SMOTE LR

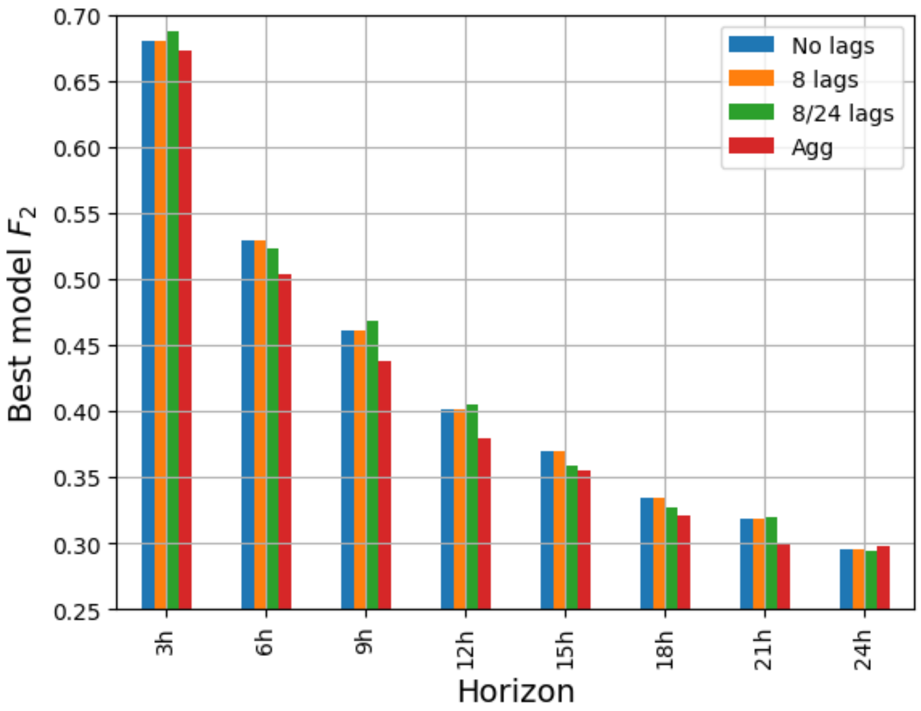
Conclusions

- **SMOTE** is key factor of model performance (especially on last horizons);
- The best model on all horizons is the **Logistic Regression;**
- Lags improve quality of the best model;
- Time series CV improve quality of the best model on several horizons;
- Not much difference between lags frequency (3h vs 1h).

Data preprocessing techniques comparison

- Increase in quality with lags is small;
- Difference between 3h and 1h lags is not significant.

Best score for each horizon and each preprocessing method



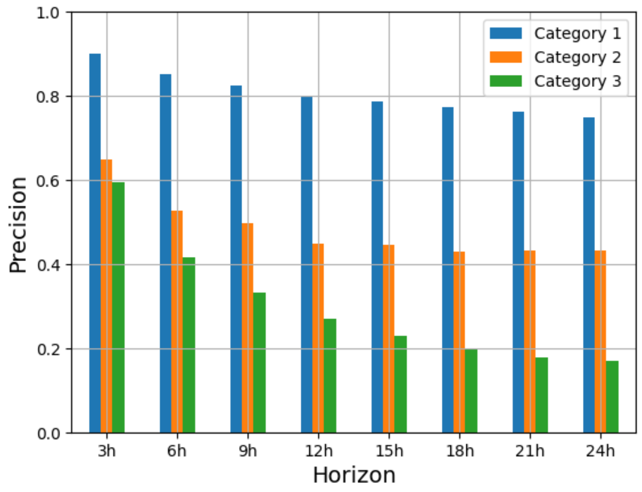
The best model's perfomance

Below is precision/recall for each category on each horizon for the best model.

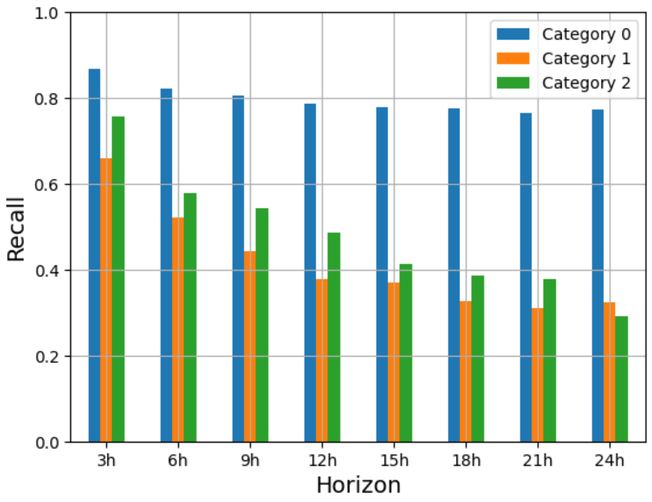
For example, **recall 75.7% and precision 59.3%** on 3 hour horizon

(vs 54.17% and 54.17% for the trivial model)

Precision for each horizon



Recall for each horizon



The best model demo (Sep-2022 –Dec-2022)

Forecast 3h ahead

Forecast 6h ahead

Forecast 9h ahead

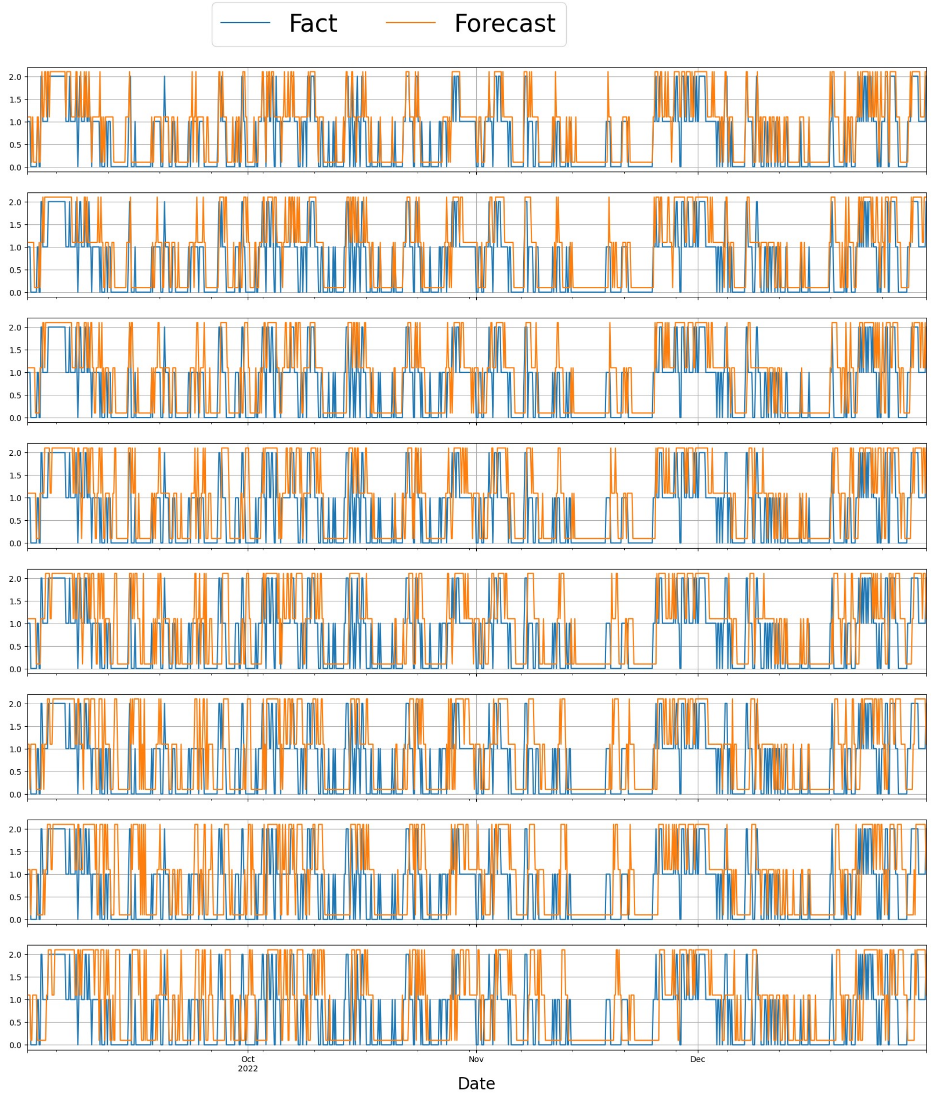
Forecast 12h ahead

Forecast 15h ahead

Forecast 18h ahead

Forecast 21h ahead

Forecast 24h ahead



Feature importances
(from 8/24 lags SMOTE LR) for category 3

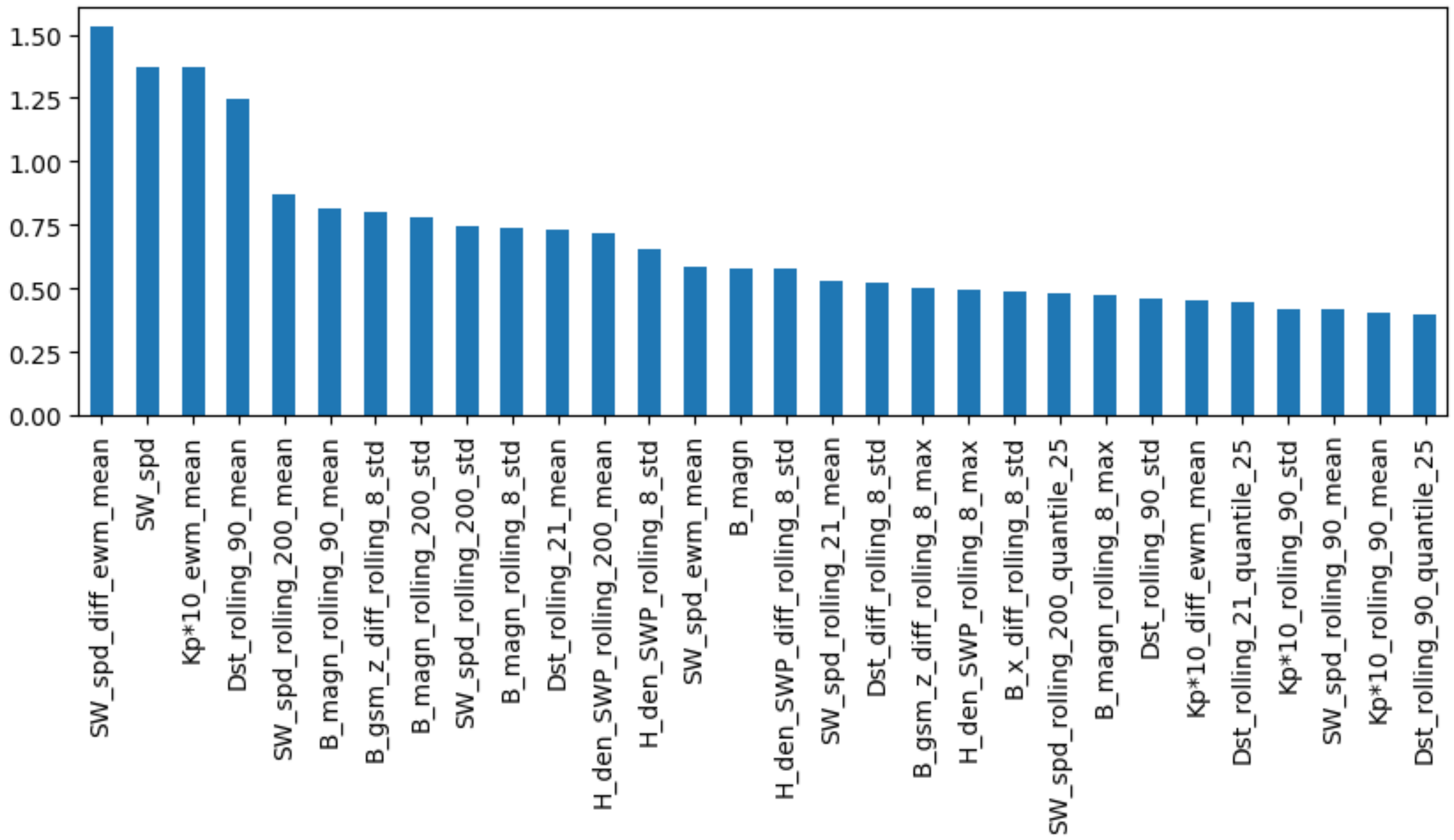
	Kp*10	Dst	B_x	B_gsm_y	B_gsm_z	B_magn	SW_spd	H_den_SWP	doySin	hourSin	doyCos	hourCos
0	1.553	-1.325	-0.066	0.020	-1.228	1.724	2.628	0.893	0.000	0.006	0.000	0.010
1		0.614	0.083	0.066	-0.395	-0.188	-0.599	-0.013	0.000	0.003	0.000	0.011
2		0.464	-0.066	-0.022	0.175	-0.156	-0.166	-0.021	0.000	-0.000	0.000	0.011
3	0.220	0.002	0.025	0.029	0.234	-0.062	-0.467	-0.163	0.000	-0.003	0.000	0.011
4		-0.164	0.042	0.050	0.054	-0.179	-0.409	-0.071	0.000	-0.006	0.000	0.010
5		-0.006	-0.070	-0.063	0.104	0.037	0.369	0.034	0.000	-0.008	0.000	0.008
6	0.178	0.177	-0.051	-0.025	0.038	-0.120	-0.445	-0.043	0.000	-0.010	0.000	0.006
7		-0.172	0.033	0.038	0.107	0.016	-0.060	-0.054	0.000	-0.011	0.000	0.003
8		0.099	0.009	0.021	-0.043	-0.117	0.140	-0.011	0.000	-0.011	0.000	-0.000
9	0.150	-0.078	0.042	-0.070	-0.011	0.113	0.010	-0.001	0.000	-0.011	0.000	-0.003
10		0.259	-0.028	-0.010	0.027	-0.205	-0.171	-0.043	0.000	-0.010	0.000	-0.006
11		-0.101	-0.020	-0.019	0.043	-0.050	-0.032	-0.053	0.000	-0.008	0.000	-0.008
12	0.098	-0.188	-0.068	0.032	0.015	0.174	-0.151	0.061	0.000	-0.006	0.000	-0.010
13		0.217	0.052	0.009	0.003	0.062	0.063	-0.025	0.000	-0.003	0.000	-0.011
14		0.016	0.062	-0.063	0.034	-0.153	-0.049	-0.044	0.000	0.000	0.000	-0.011
15	0.082	0.082	0.022	0.083	-0.044	0.066	0.319	-0.014	0.000	0.003	0.000	-0.011
16		0.012	-0.044	-0.052	0.062	0.030	-0.232	-0.045	0.000	0.006	0.000	-0.010
17		-0.077	0.009	0.011	-0.056	-0.027	-0.031	-0.011	0.000	0.008	0.000	-0.008
18	0.119	-0.080	-0.043	0.013	0.032	-0.065	0.170	-0.017	0.000	0.010	0.000	-0.006
19		0.211	-0.009	0.059	-0.013	0.032	-0.208	-0.023	0.000	0.011	0.000	-0.003
20		-0.332	0.012	-0.050	0.011	0.035	-0.067	0.060	0.000	0.011	0.000	0.000
21	0.150	0.256	0.005	-0.015	0.042	-0.017	-0.200	-0.056	0.000	0.011	0.000	0.003
22		-0.095	0.025	-0.036	0.032	-0.137	0.429	0.000	0.000	0.010	0.000	0.006
23		0.186	-0.021	0.078	-0.008	0.158	-0.269	0.003	0.000	0.008	0.000	0.008
24	0.036	-0.136	-0.010	-0.051	0.008	-0.184	0.088	-0.022	0.001	0.006	0.000	0.010

2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0

Feature importances (from 8/24 lags SMOTE LR) for category 3

- The most important feature is the current value of solar wind speed **SW_spd** (and its lags)
- Top 5 features (by module):
  - SW_spd (+2.62)
  - B_magn (1.72)
  - Kp*10 (1.55)
  - Dst (-1.32)
  - B_gsm_z (-1.228)

Feature importances (from AGG SMOTE LR) for category 3



Thank you for your attention!