REPLY
DATA

AI FOR ENVIRONMENT-PREDICTIVE MAINTENANCE ON WIND TURBINE DATA

MAIN CHALLENGE

26th of April 2024

PREDICTIVE MAINTENANCE ON WIND TURBINES DATA

Problem Statement

Develop a solution to **optimize** wind turbine **performance** and maintenance scheduling for **increased energy efficiency** and **operational reliability.**

To ensure the profitability of wind turbines, it is essential to develop predictive maintenance solutions that will:

- Optimize energy production (early warnings & improve equipment efficiency)
- Prevent unexpected downtimes (decrease system downtime of wind turbine operation)

The goal is to leverage data from wind turbine systems to minimize downtime, maximize power output, and extend the lifespan of wind turbines.

HOW DO YOU OPTIMIZE ENERGY PRODUCTION AND PREVENT UNEXPECTED DOWNTIMES?

Our solution:

Predict **specific technical problems** that **caused** the windmill to **stop operating** multiple times in the past.

These **errors** need to have some **properties**:
- Need to happen **frequently** and idealy on all operating machines
- Can be **predicted** from the data
- Have to be **preventable** by maintanance work in advance (e.g. strong wind that causes reduced RPM is not relevant)

→Problem: The given dataset does not provide this information.

So we searched for our own dataset first: Penmanshiel Wind Farm Data (GB), Kelmarsh Wind Farm (GB) with **299 features and more than 1.7 mio. time stamps**

HOW DO YOU OPTIMIZE ENERGY PRODUCTION AND PREVENT UNEXPECTED DOWNTIMES?

1. **Understanding Failures** frequency, predictability, and preventability, focused on frequency converter errors
2. **Feature Reduction** only less than 18% NaN, determine 48 relevant attributes
3. **Error Data Analysis** Quantified the occurrence of frequency converter errors and the associated downtime
4. **Feature Data Analysis** visual inspections to detect patterns and correlations
5. **Visualization and Dimension Reduction with Umap** Detected clusters, multiple counts of events per day
6. **Data Cleaning / Insufficient Data for Reliable Predictions** only 14 points available for predictive modeling.
7. **Time Series Analysis** intervals of 24, 48, and 72 hours before damage, aggregated features
8. **Further Feature Reduction and Predictive Modeling** feature importance metrics, logistic regression model
9. **Model Testing** developed model to the Penmanshiel, did not yield successful predictions, data limitations

1.UNDERSTANDING FAILURES

[21]

```
error_labels["date"] = error_labels["Timestamp start"].dt.date

error_labels = error_labels.drop_duplicates(subset=['Turbine','Message','date'])
```

[22]

```
error_labels
```

...

	index	Timestamp start	Turbine	Message	date
0	70	2016-01-29 10:08:33	2	Frequency converter error	2016-01-29
1	77	2016-01-30 07:33:38	2	Frequency converter error	2016-01-30
2	81	2016-02-01 18:43:27	2	Frequency converter error	2016-02-01
3	83	2016-02-02 16:04:36	2	Frequency converter error	2016-02-02
5	293	2016-03-15 12:51:40	2	Overload generator fan 1	2016-03-15
...	...	...	...	...	...
385	271763	2021-01-19 10:38:48	5	Overload generator fan 2	2021-01-19
388	280499	2021-05-15 06:55:24	4	Overload generator fan 1	2021-05-15
389	280603	2021-05-16 07:59:48	4	Overload generator fan 2	2021-05-16
390	280604	2021-05-16 07:59:49	4	Overload generator fan 3	2021-05-16
391	280605	2021-05-16 07:59:50	4	Overload generator fan 1	2021-05-16

focus on one error

273 rows × 5 columns

2.FEATURE REDUCTION

#	Attribute Name	Attribute Description
1	Long Term Wind (m/s)	Average wind speed over a longer period of time
2	Grid frequency (Hz)	The frequency of the electrical grid
3	Rotor speed (RPM)	revolutions per minute
4	Generator RPM (RPM)	The rotational speed of the generator within the wind turbine
5	Hub temperature (°C)	The temperature at the hub of the wind turbine
6	Gear oil temperature (°C)	The temperature of the gear oil within the wind turbine gearbox
7	Gearbox speed (RPM)	The rotational speed of the gearbox within the wind turbine
8	Transformer temperature (°C)	The temperature of the transformer used to convert electrical energy
9	Energy Export (kWh)	The amount of electrical energy exported or generated by the wind turbine

3.ERROR DATA ANALYSIS

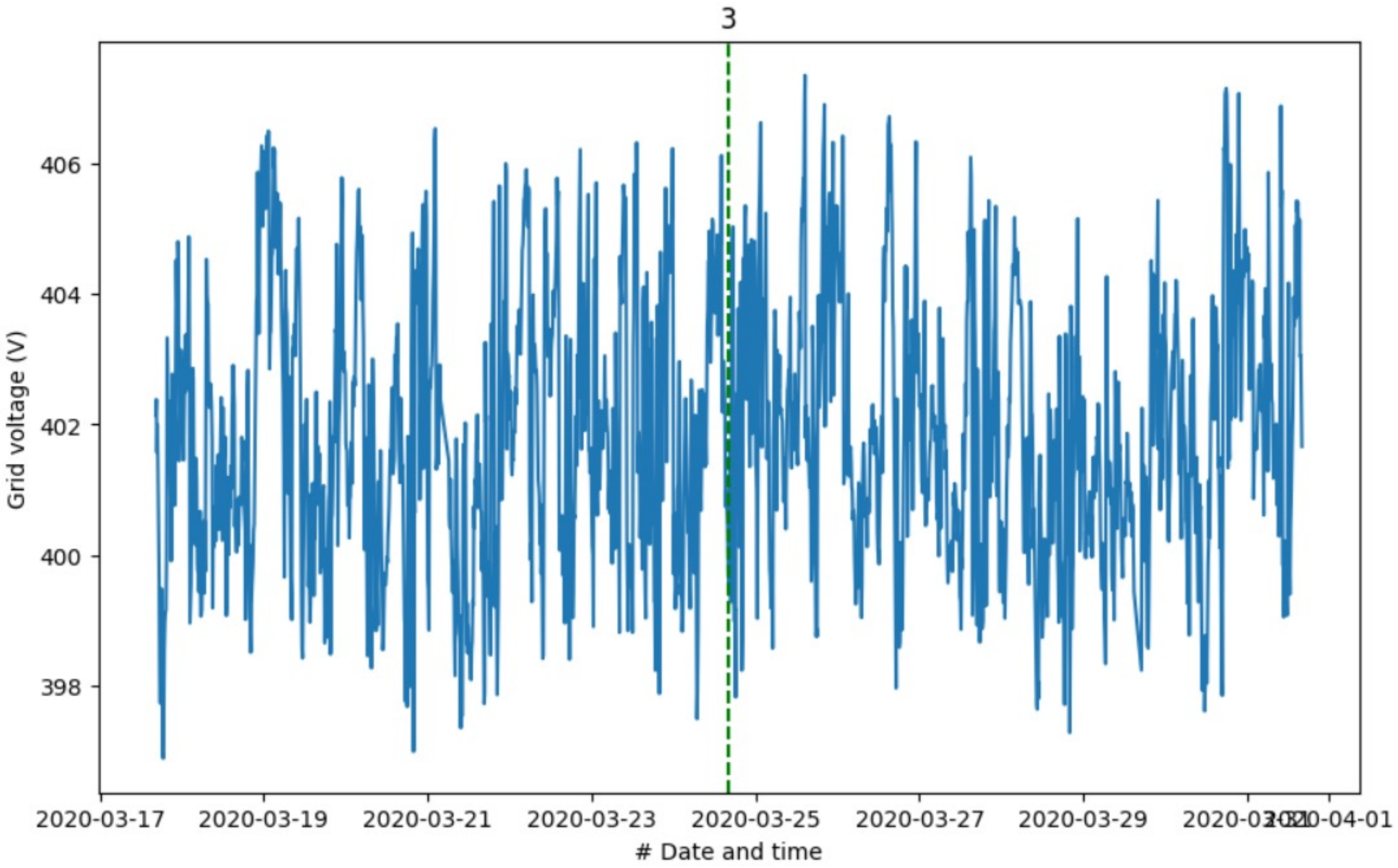
```python
for error in sorted_errors:
    selector = status_reduced.Message == error[0]
    total_duration = status_reduced[selector].duration_timedelta.sum()

    print(error[0],"---",error[1],"---",total_duration)
```

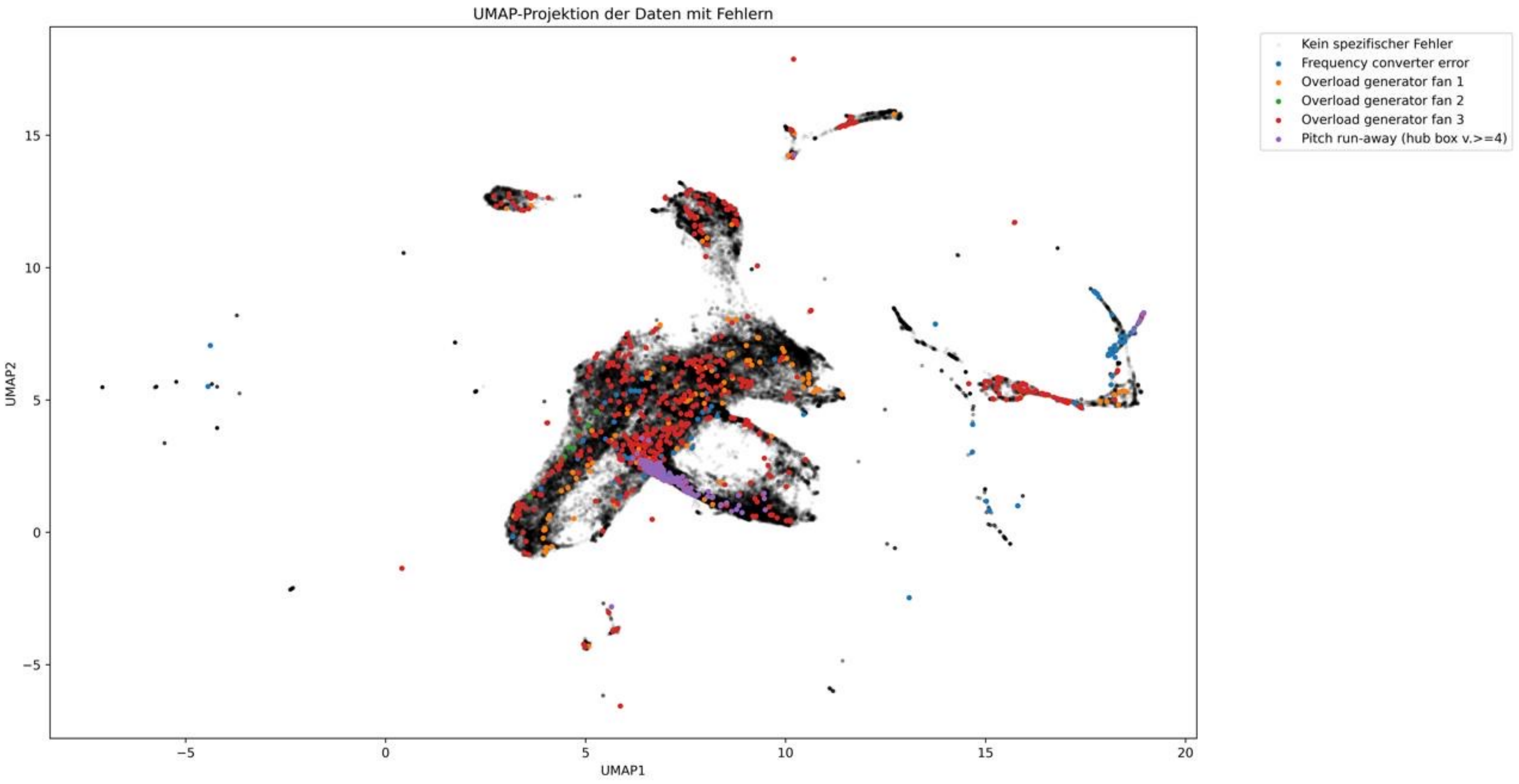
```
Overload generator fan 1 --- 81 --- 21 days 00:10:32
Overload generator fan 2 --- 67 --- 19 days 22:23:16
Overload generator fan 3 --- 66 --- 19 days 19:34:08
Manual stop - remote --- 23 --- 10 days 08:17:23
Frequency converter error --- 19 --- 30 days 22:45:16
Safety chain open --- 17 --- 5 days 18:41:27
Frequency converter not ready --- 17 --- 1 days 09:33:45
Limit switch error 95° axis 1 --- 15 --- 1 days 07:56:52
High rotor speed nacelle --- 13 --- 1 days 22:53:01
Repeating error BP52 --- 12 --- 5 days 08:31:42
Pitch run-away (hub box v.>=4) --- 10 --- 0 days 06:06:01
Uncontrolled yaw movement --- 8 --- 0 days 00:42:13
Oscillation encoder tower --- 7 --- 1 days 16:59:40
Yaw speed high --- 5 --- 0 days 00:28:31
Maximum grid frequency --- 5 --- 2 days 19:00:03
Low hydraulic pressure --- 5 --- 2 days 04:11:40
Emergency stop base box --- 4 --- 0 days 04:32:50
Low gearbox oil pressure --- 4 --- 4 days 02:22:16
Emergency stop top box --- 3 --- 0 days 03:13:14
Pitch angle deviation --- 3 --- 0 days 06:17:26
Feedback brake 1 --- 3 --- 2 days 09:19:21
```

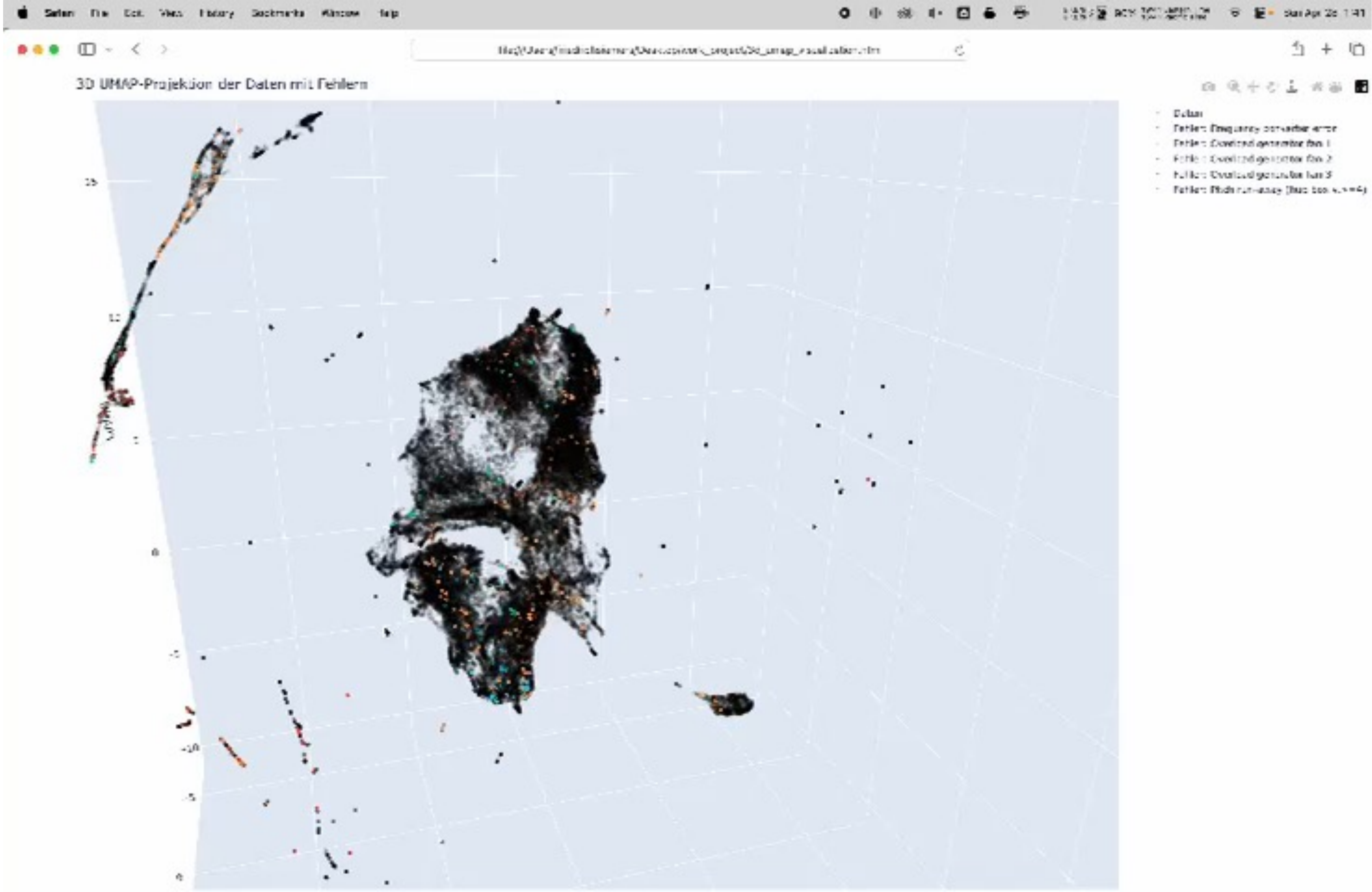
more than 30 days of downtime on the data in Kelmarsh

4.FEATURE DATA ANALYSIS



5. Visualization and Dimension Reduction with Umap




3D UMAP-Projektion der Daten mit Fehlern

6.DATA CLEANING / INSUFFICIENT DATA FOR RELIABLE PREDICTIONS

```
df = error_labels
# Assuming 'df' is your DataFrame
# First, make sure the 'date' column is in datetime format
df['date'] = pd.to_datetime(df['date'])

# Sort the DataFrame by 'Turbine' and 'date'
df = df.sort_values(by=['Turbine', 'date', 'Message'])


# Check for consecutive messages for the same turbine on consecutive days
df['duplicate'] = df.duplicated(subset=['Turbine', 'Message'], keep='first') & \
                  (df['date'] - df['date'].shift() == pd.Timedelta(days=1)) & \
                  (df['Turbine'] == df['Turbine'].shift())
```

3]

```
# Keep only the non-duplicate entries
df = df[~df['duplicate']]

# Drop the 'duplicate' column as it's no longer needed
df.drop(columns=['duplicate'], inplace=True)

# Reset the index of the DataFrame if needed
df.reset_index(drop=True, inplace=True)
```

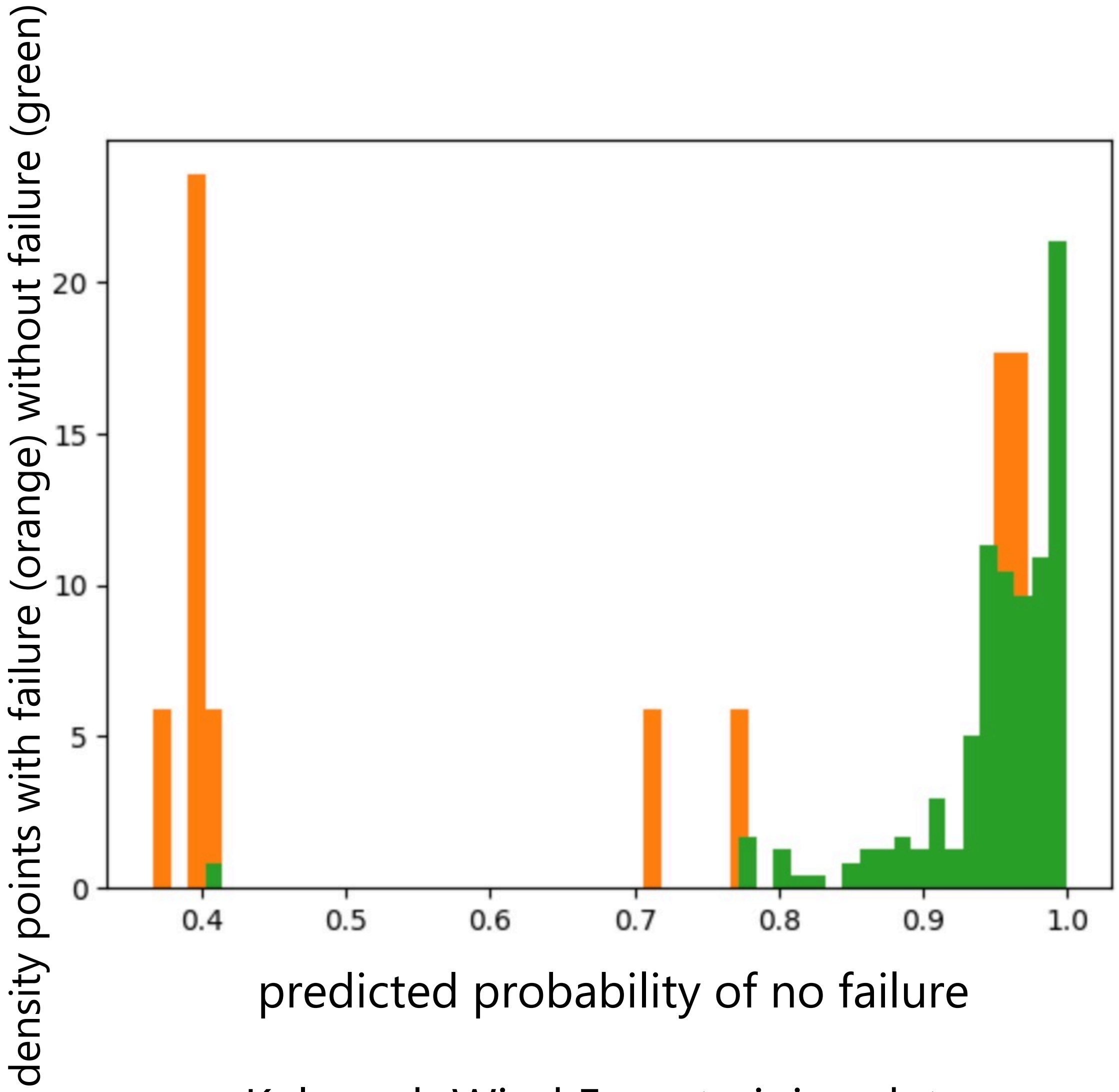
46]

	index	Timestamp start	Turbine	date
0	10012	2016-01-24 16:51:17	1	2016-01-24
1	10157	2016-03-01 18:35:40	1	2016-03-01
2	70	2016-01-29 10:08:33	2	2016-01-29
3	81	2016-02-01 18:43:27	2	2016-02-01
4	64048	2018-03-14 10:50:34	2	2018-03-14
5	72499	2018-12-17 15:40:51	2	2018-12-17
6	1892	2016-02-14 11:12:49	3	2016-02-14
7	90818	2018-04-20 05:31:19	3	2018-04-20
8	230649	2020-10-24 12:09:21	3	2020-10-24
9	132687	2018-11-02 17:58:41	4	2018-11-02
10	198975	2020-08-21 07:00:30	4	2020-08-21
11	7920	2016-02-04 20:03:42	5	2016-02-04
12	4730	2016-02-22 13:18:07	6	2016-02-22
13	46264	2017-11-11 20:05:53	6	2017-11-11

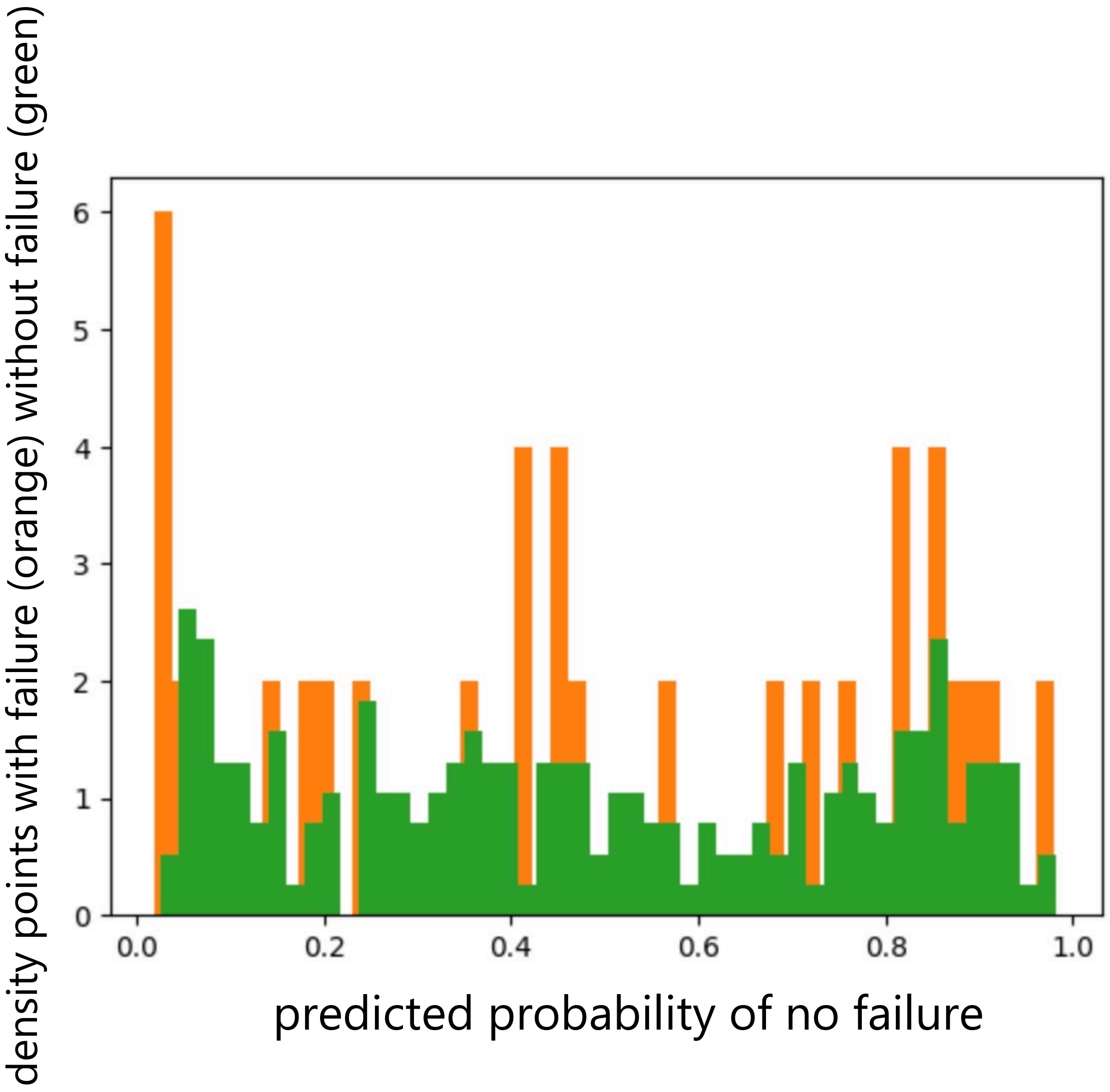
7. **Time Series Analysis** intervals of 24, 48, and 72 hours before damage, aggregated features

8. **Further Feature Reduction and Predictive Modeling** feature importance metrics, logistic regression model

9.MODEL TESTING



Kelmarsh Wind Farm training data



Penmanshiel Wind Farm test data

did not yield successful predictions due to data limitations

FUTURE DIRECTIONS

Expand data collection efforts to include a wider array of error types and operational conditions.

Explore **more advanced machine learning algorithms** that can handle sparse data more effectively.

REPLY
DATA

THANK YOU FOR YOUR TIME

April 2024, Munich.

CONTACT

Phillip Wondra:
www.linkedin.com/in/philipp-wondra-76552719b

Maximilian Seeger

Benedikt Kellner

Friedrich Siemers

DATA SOURCE:

https://zenodo.org/records/5841834#.YgpBQ_so-V7
https://zenodo.org/records/5946808#.YgpAmvso-V5

Friedrich Siemers

Entrepreneurial Polymath AI, Math, Computational Linguistics & Philosophy Founder of KImpact, Generating Measurable Impact for Small and Medium-Sized Companies

www.kimpact.de