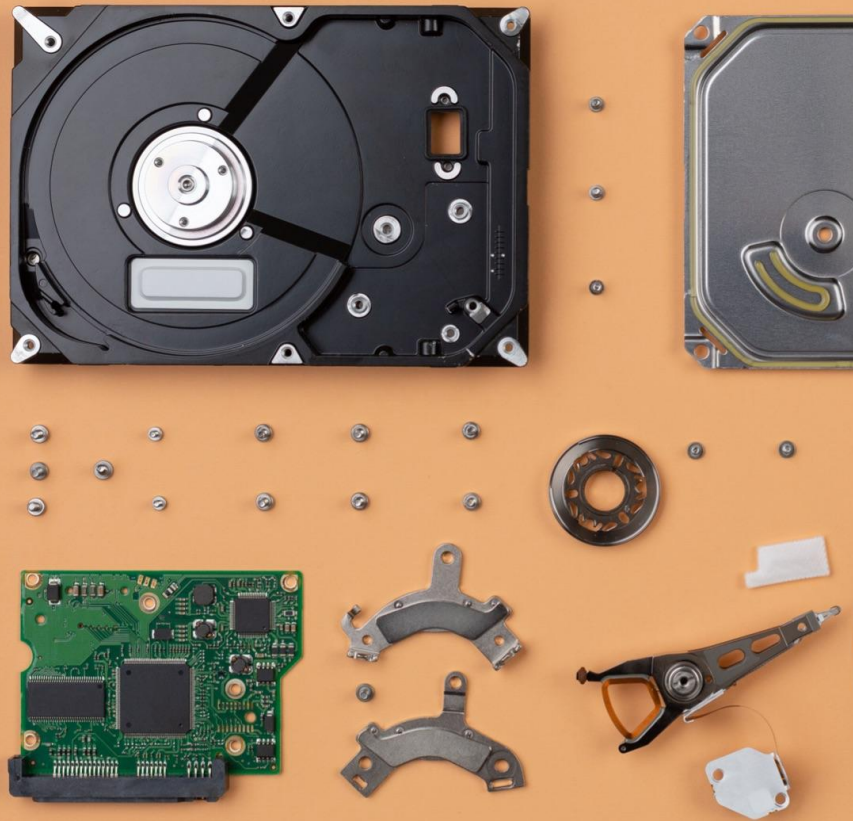


# To Fail or not to Fail?

Predicting Hard Drive  
Health

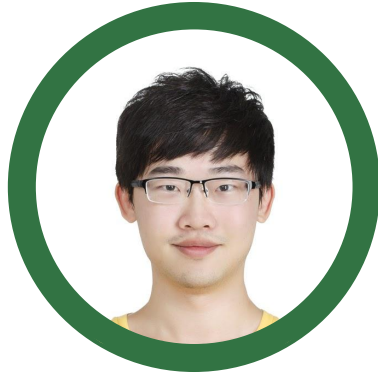


# Guardians of the Memory



**Andreas**

Dipl.-Ing.  
mech. Engineering  
with a background in  
energy technology



**Chang-Ming**

PhD in Physics with  
a background in  
theoretical modeling



**Daniela**

Application  
Manager with a  
background in  
language studies

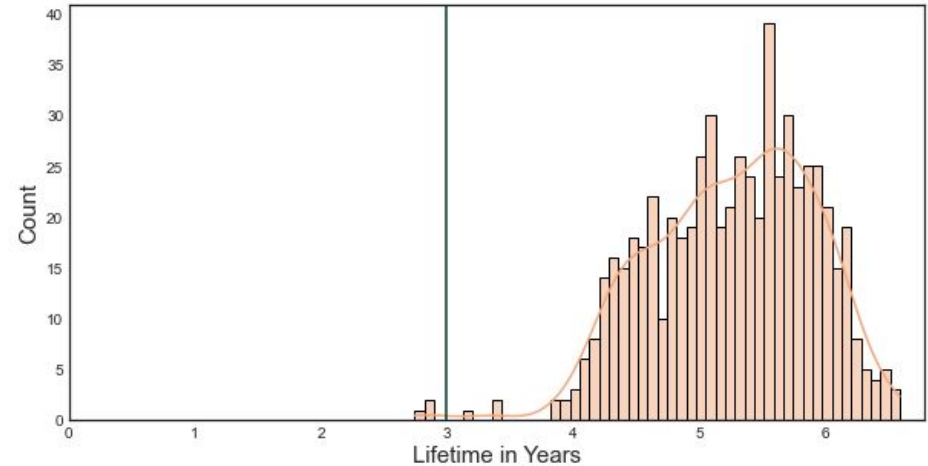


**Felix**

PhD in Physics with  
strong background  
in data analytics

# The Stakeholder - Cloudwaver

- Startup offering cloud storage as a service
- Maximize hard drive usage beyond 3 years



# The Task

**Predict if a hard drive fails in the coming 30 days:**

- **Reduce investments** for hard drives by up to **40%**  
(5 years vs 3 years of usage)
- **Enhance sustainability**
- Maintain growth even under global chip shortage



# The Hard Drives Dataset

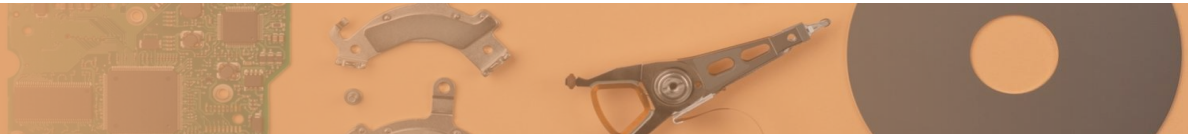
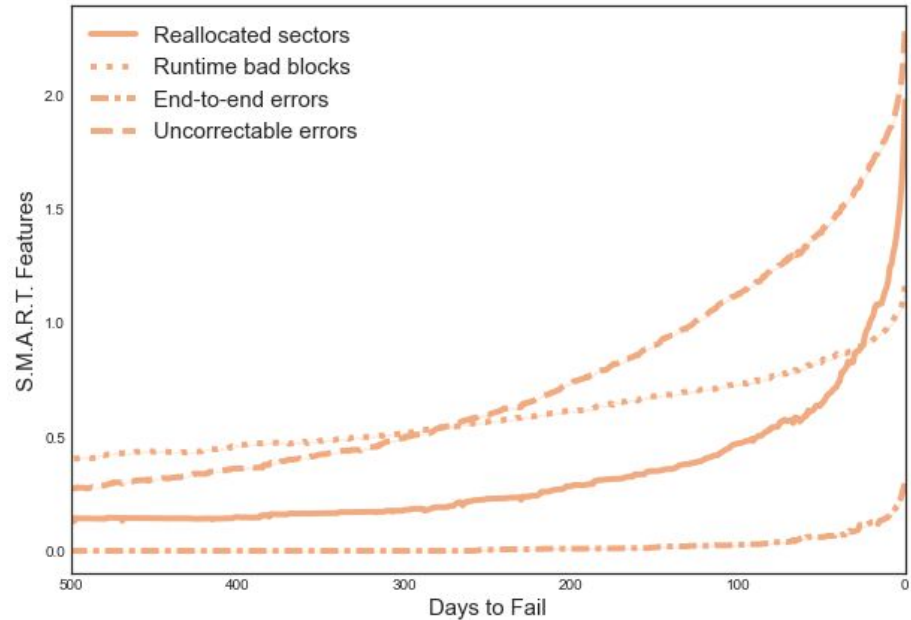


- Around 205k hard drives in 65 models (2021)
- 174 S.M.A.R.T. parameters recorded daily
- **S.M.A.R.T:** Self-Monitoring, Analysis and Reporting Technology
- Focus on the model of interest for Cloudwaver (2019 to 2021)

# Crucial S.M.A.R.T. Features

Examples of S.M.A.R.T. Features:

- **Reallocated sectors**
- **Runtime bad blocks**
- **End-to-end errors**
- **Uncorrectable errors**
- Temperature
- Power on time



# Model Evaluation

**F2-score: accounts for recall and precision**

**Healthy**

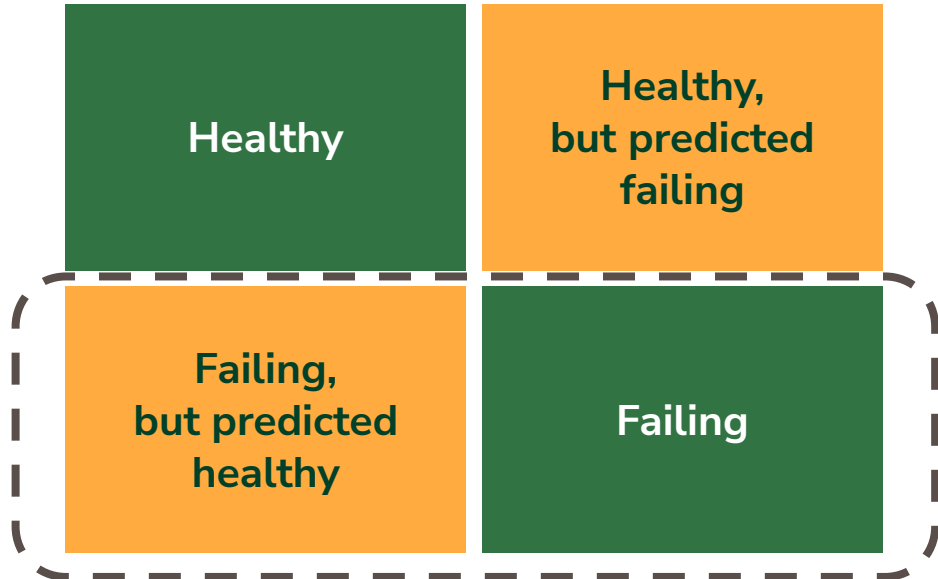
**Healthy,  
but predicted  
failing**

**Failing,  
but predicted  
healthy**

**Failing**

# Model Evaluation

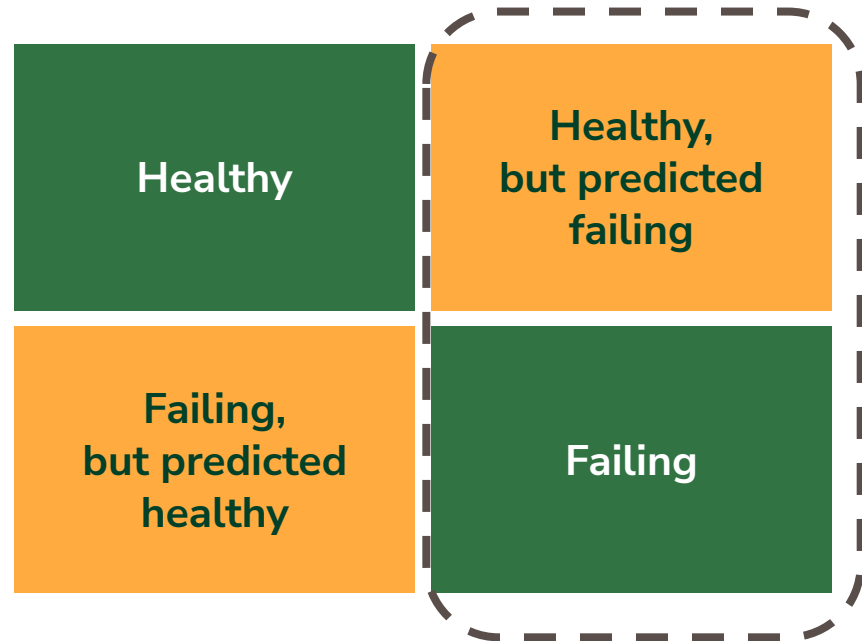
F2-score: accounts for recall and precision



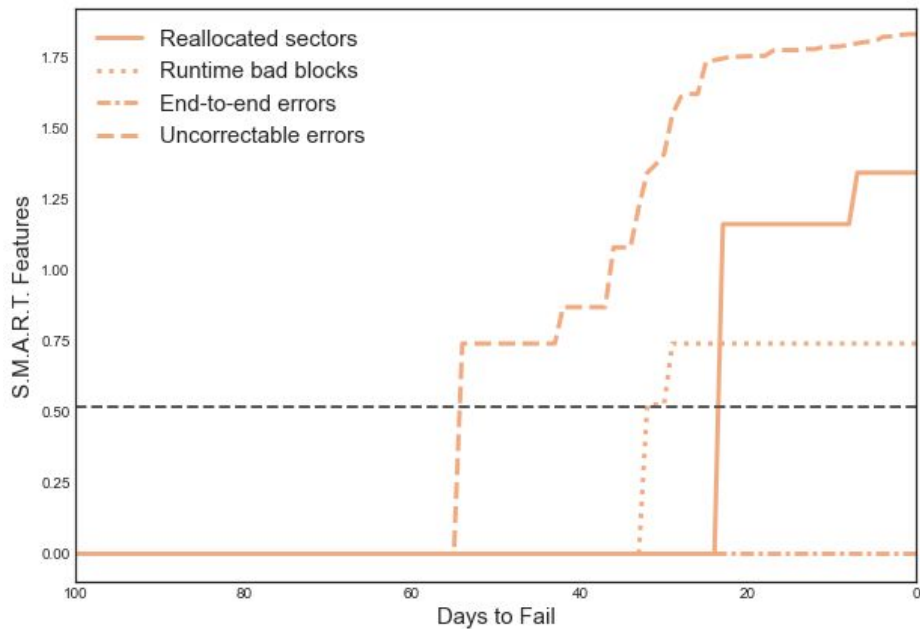


# Model Evaluation

F2-score: accounts for recall and precision



# Baseline Model



**29 %**

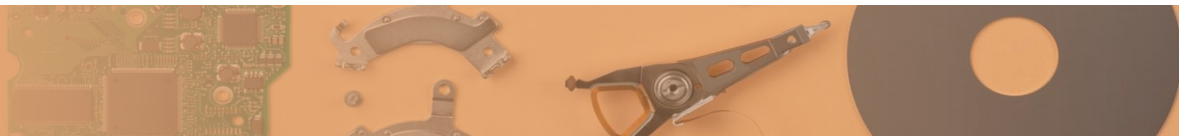
F2-Score

**45 %**

Recall

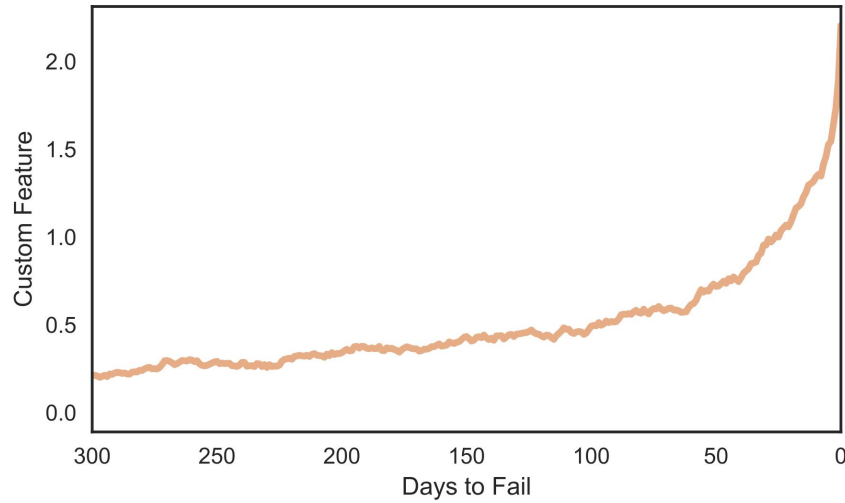
**12 %**

Precision



# Final Model

- Low-dimensional artificial neural network
- Custom feature captures dynamics of relevant S.M.A.R.T. features



**44 %**

F2-Score

**61 %**

Recall

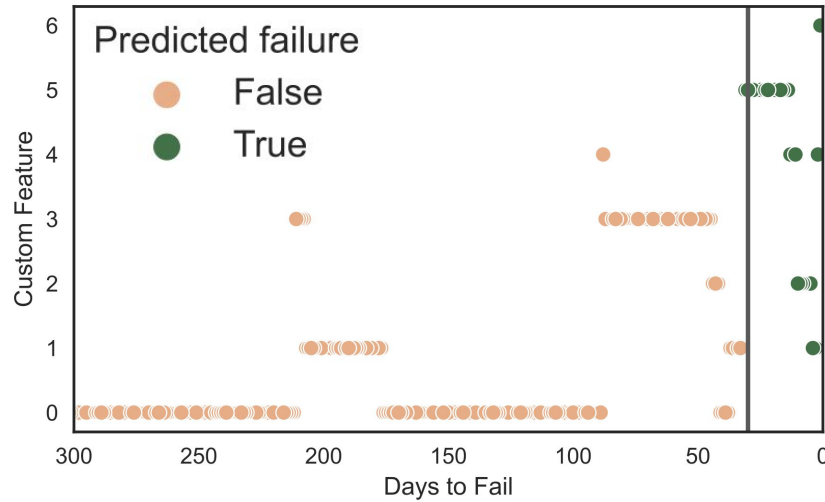
**21 %**

Precision



# Final Model

- Low-dimensional artificial neural network
- Custom feature captures dynamics of relevant S.M.A.R.T. features



44 %

F2-Score

61 %

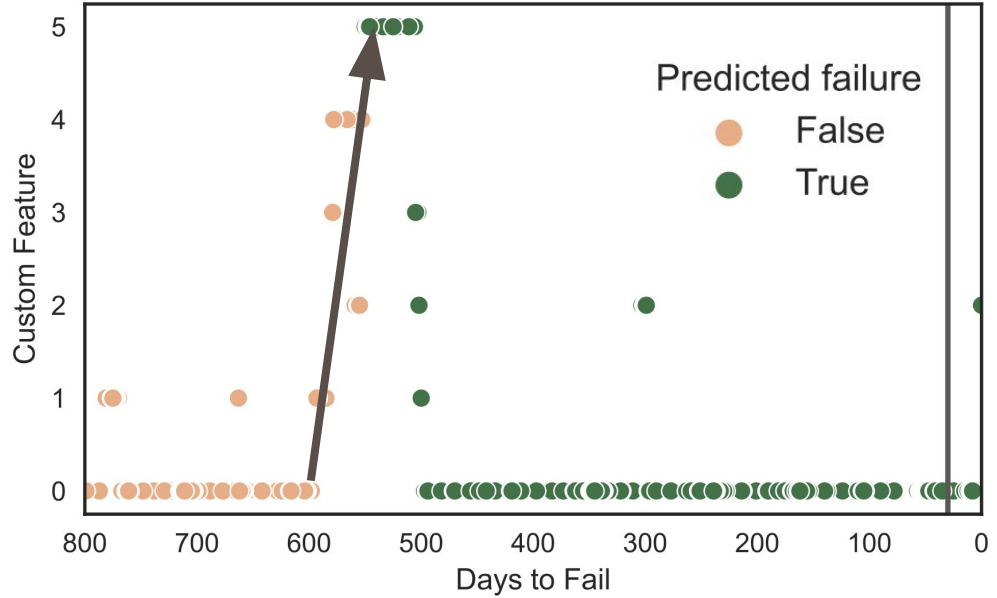
Recall

21 %

Precision

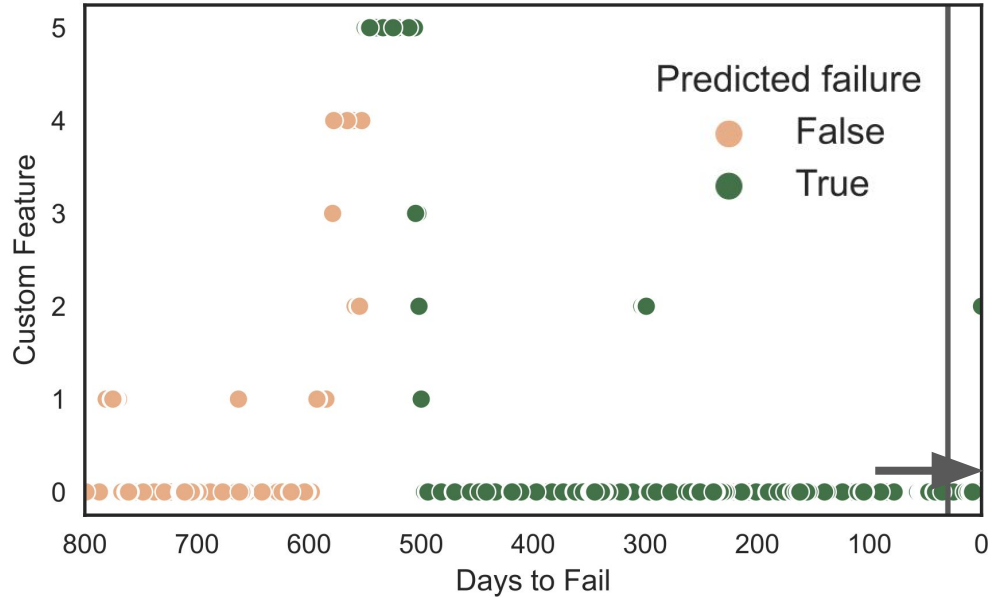
# Limitations

- “Fake” failures



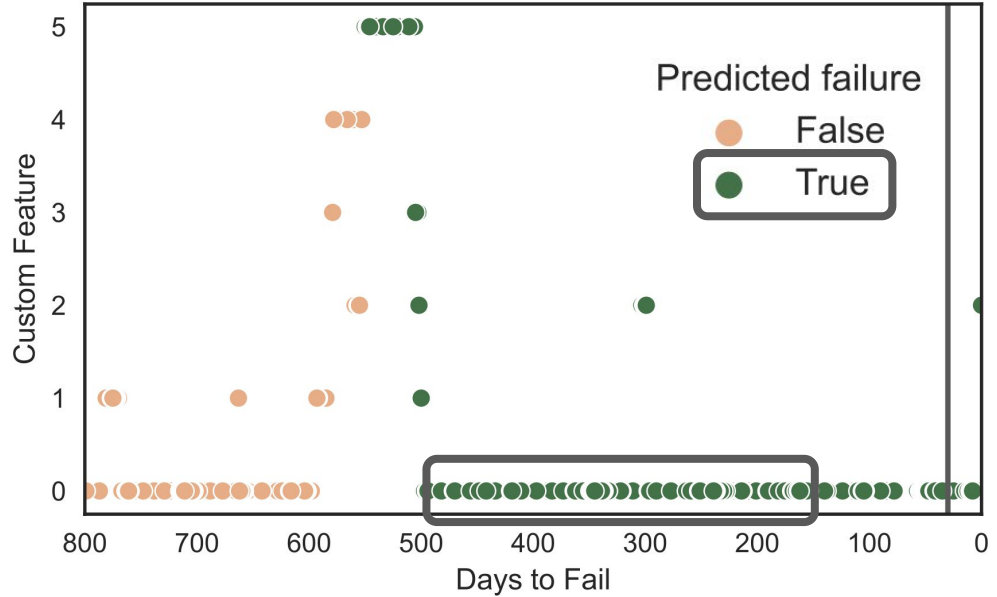
# Limitations

- “Fake” failures
- “Silent” failures



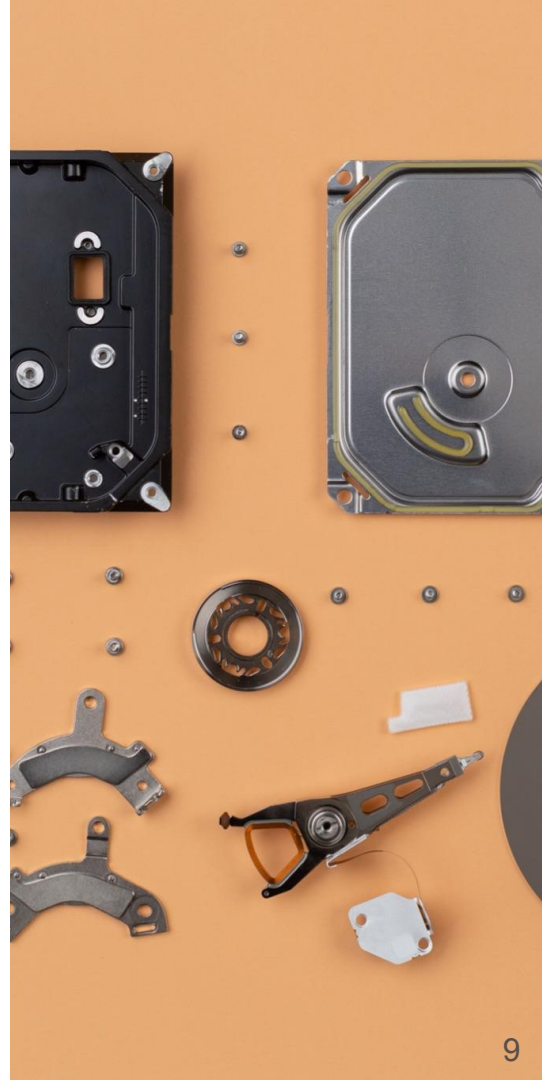
# Limitations

- “Fake” failures
- “Silent” failures
- Time-related features



# Outlook

- Try to tackle limitations
- Anomaly detection methods
- Include additional hard drive models





Fail or not to Fail!



Guardians of the Memory

Felix, Chang Ming, Andreas & Daniela



# How long will your hard drive last?

This is a web app to predict if a HDD drive will fail or not fail in the next 30 days. Please click on the Predict button to see the results of the classification.

This is how a random sample of our raw data looks like:

	date	serial_number	model	capacity_bytes	failure	smart_1_no
58	2021-03-31	Z3058TQY	ST4000DM000	4000787030016	0	
59	2021-03-30	Z3058TQY	ST4000DM000	4000787030016	0	
60	2021-03-29	Z3058TQY	ST4000DM000	4000787030016	0	
61	2021-03-28	Z3058TQY	ST4000DM000	4000787030016	0	
62	2021-03-27	Z3058TQY	ST4000DM000	4000787030016	0	

Predict on our provided test data

