Taking your data for a spin!

S.M.A.R.T ways of predicting hard drive failure

The Team

Kelechi John

Data Scientist / Process Engineer

Jakob Wegener

Data Scientist / Biologist

Frederik Bruns

Data Scientist / Industrial Engineer

The Stakeholder

Backblaze

- Cloud storage for business and private use
- Regular release of their hard drive statistics

The Problem

Hard drives fail

Data loss

operating costs

Self-Monitoring, Analysis and Reporting Technology (S.M.A.R.T)

history

First version introduced in 1992 by IBM

Sensor data

Log of various indicators of drive reliability

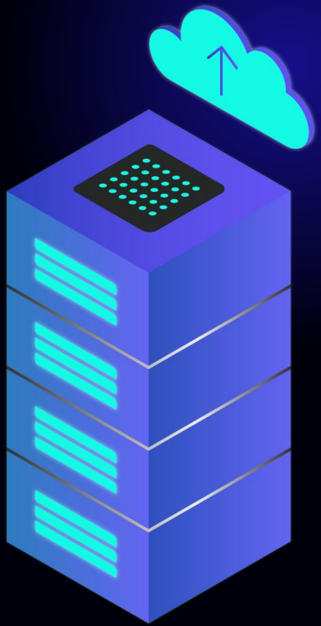
normalization

sensor readings get transformed to a scale of 0-100

limitations

Google Study*: 36% of failed drives record no S.M.A.R.T. errors

*Pinheiro et al., 2007, Failure Trends in a Large Disk Drive Population



231,309

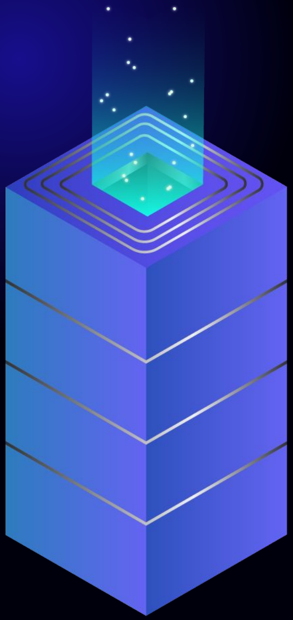
Hard Drives

178

Features

365

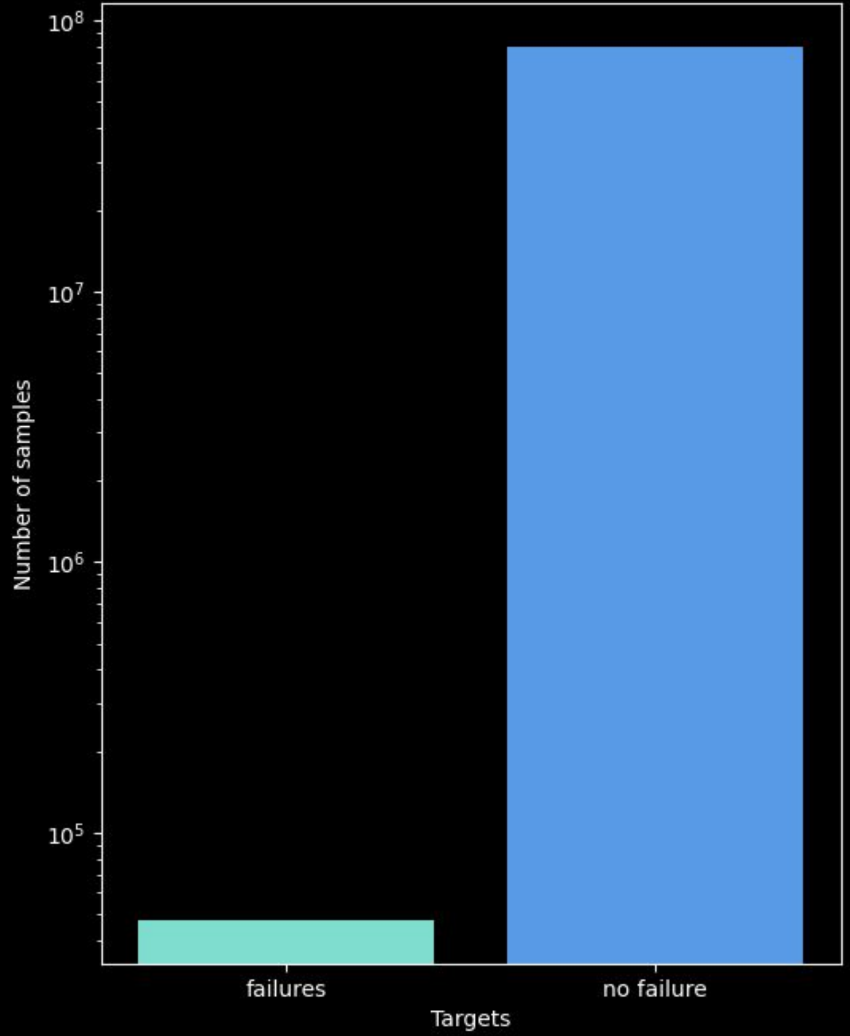
Days



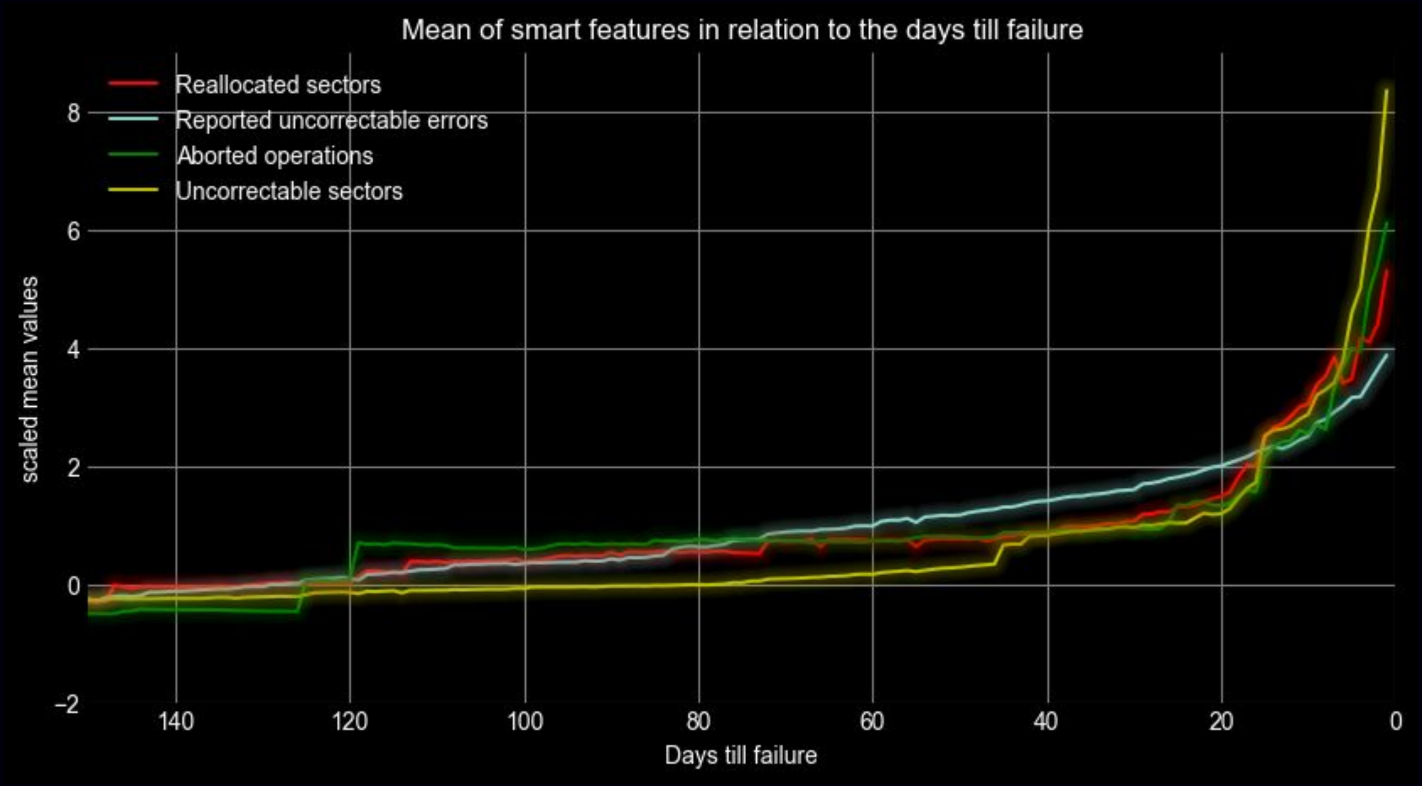
data is imbalanced!

80,357,762 observations

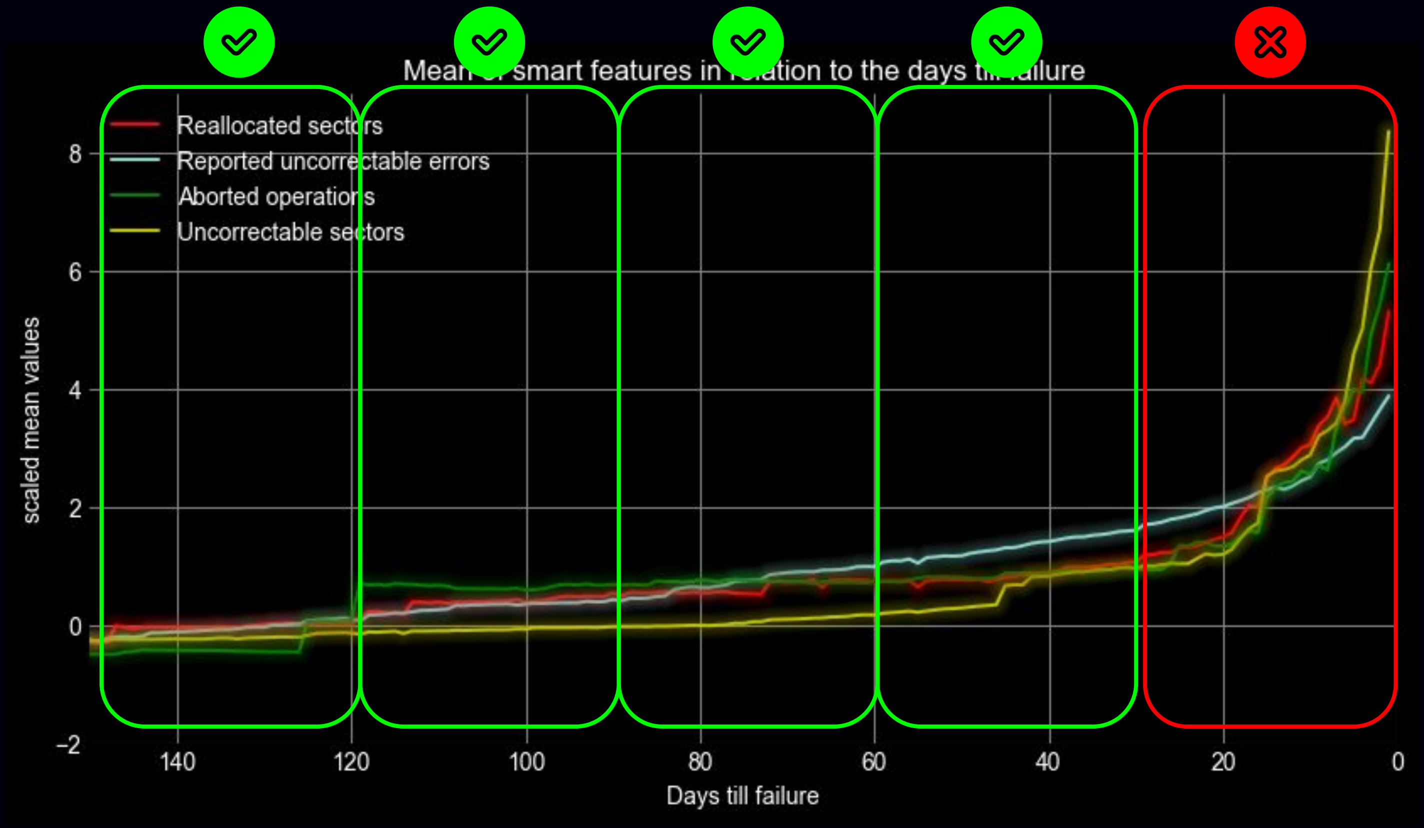
1.2% are failures and 98.8% are no failures



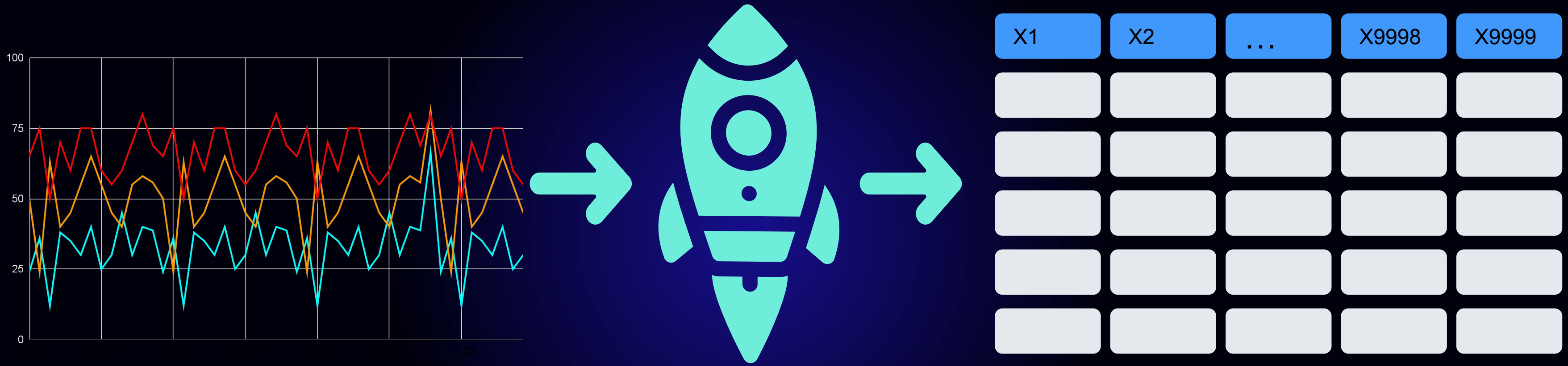
Certain smart features spike when failure is near



Split the time series into windows



How to transform time series data?



miniROCKET: A Very Fast (Almost) Deterministic Transform for Time Series Classification using convolutional kernels

Model results

Baseline model

- Logistic regression
- Based on raw data
- Ignores time as a dimension

Final Model

- Random Forest
- Based on transformed data
- Time is used in creating features

Model results

Baseline model

Accuracy 99%

Recall 3%

Precision 29%

Final Model

Accuracy 99%

Recall 57%

Precision 22%

Conclusions

Limitation of prediction

Imbalanced data
Some failures show no signal

Long term trend

30 day window does not capture failure signs before the last 30 days

Detection rate of 57%

Compared to 52%* to 60%** in literature

* Amram et al., 2021, Interpretable predictive maintenance for hard drives

** Ircio et al., 2015, A Multivariate Time Series Streaming Classifier for Predicting Hard Drive Failures

Future Outlook

Anomaly detection

Neural Network as classifier

Increase scope

THANKS!

Do you have any questions?
Join us in our breakoutroom!
Github:
https://github.com/JRJWegener/hard_drive_ML_capstone_project

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**