Electric motors data analysis

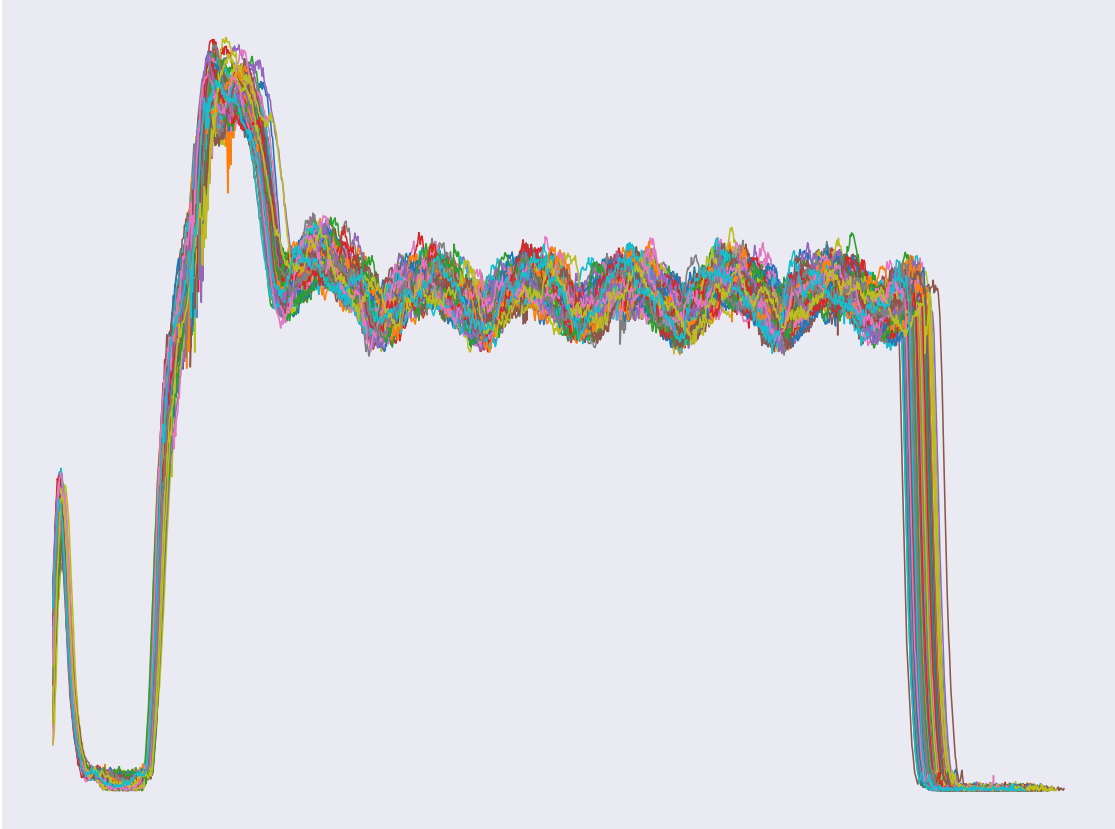
Group 8:
Viktor Snesarevskii
Edoardo Belli
Guendalina Biava
Ozrenka Dragić
Madina Kalenova

March-April 2018

Dataset

- Measurements of torque and current from electric motors
- Time-series data with frequency 20KHz (downsampled by a factor $M{=}100$ for real use cases/ease of computation)
- Data taken from 5 distinct motors in AC and DC modes
- Each data sample is one recorded operation
- Operations are recorded while the motor is working properly, then a fault is induced and the same operations are recorded again
- 1066 AC samples, 924 DC samples (focusing on the DC for now)

The Dataset



Final Goals

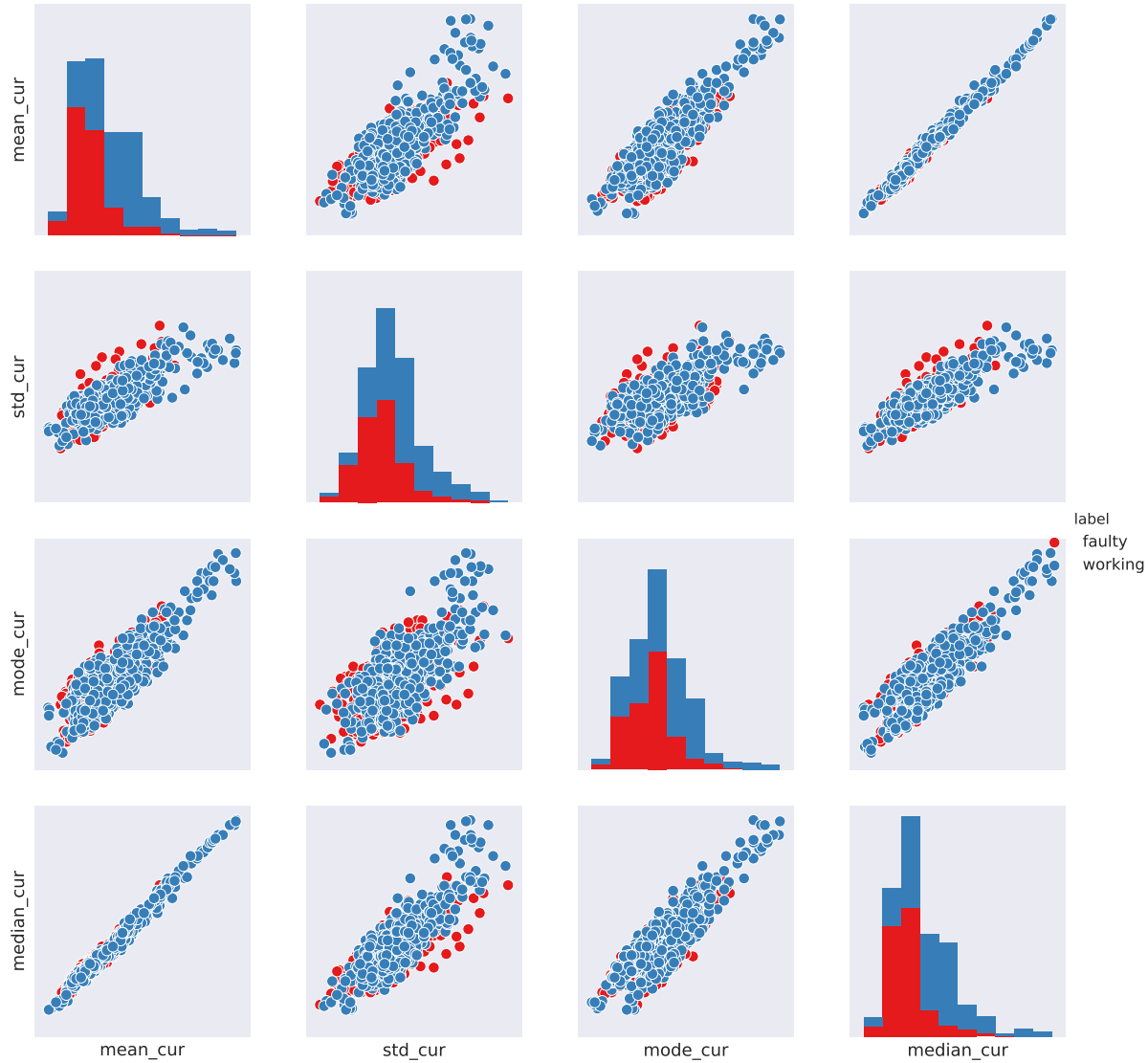
ABB is interested in three main applications:

- Motor Classification
  Can we identify a motor by a recorded operation?
- Fault Classification
  Can we group motors by their state (working/faulty)?
- Fault Prediction
  Can we predict the state of a motor by measuring current and torque during an operation?

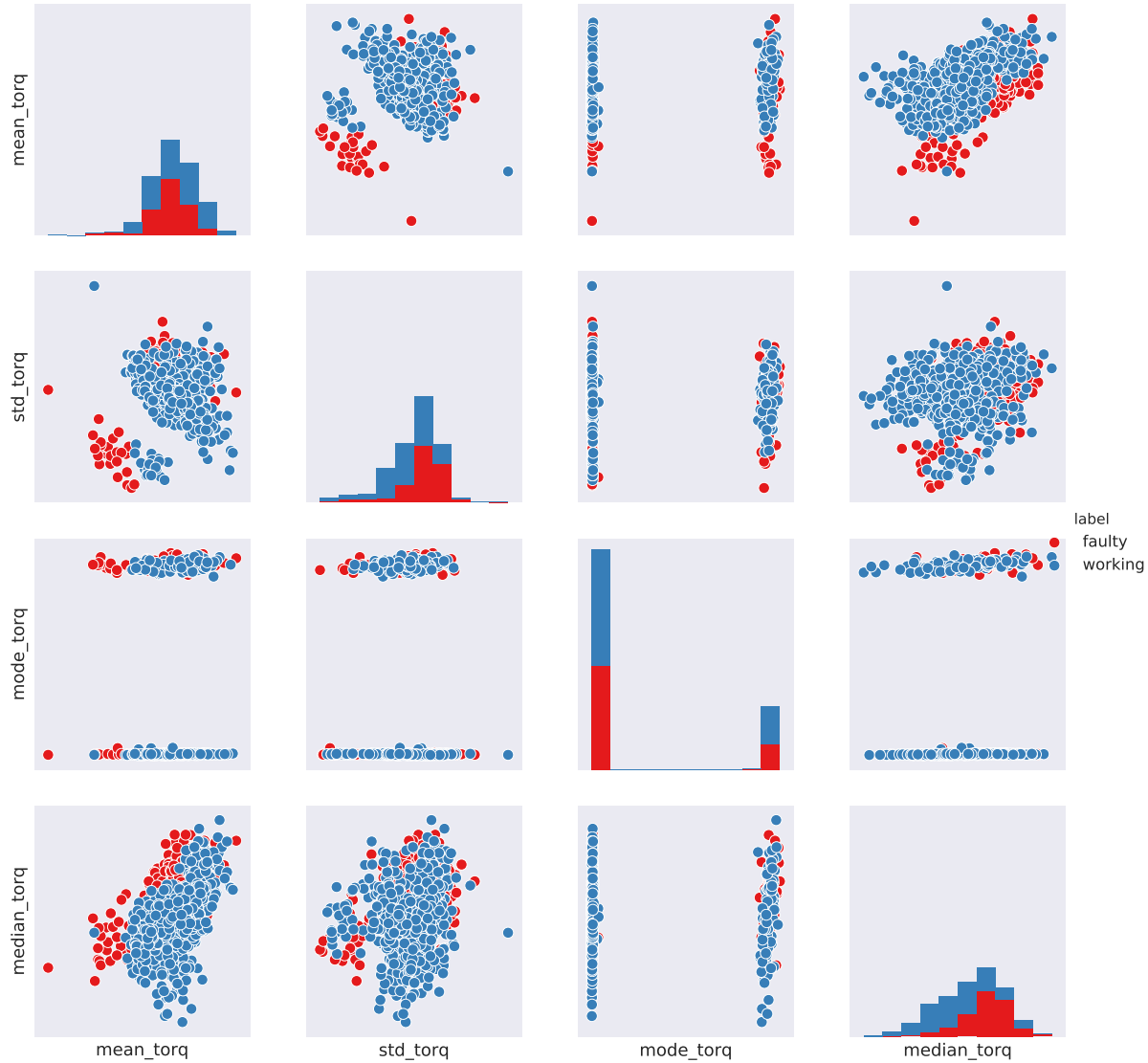
Exploratory Analysis: First Approach

- Feature Extraction
  Choose a set of features and check if they are statistically significant for the underlying problem —> to be used later in classification
- Our set of features is:
  Mean, Standard Deviation, Mode, Median

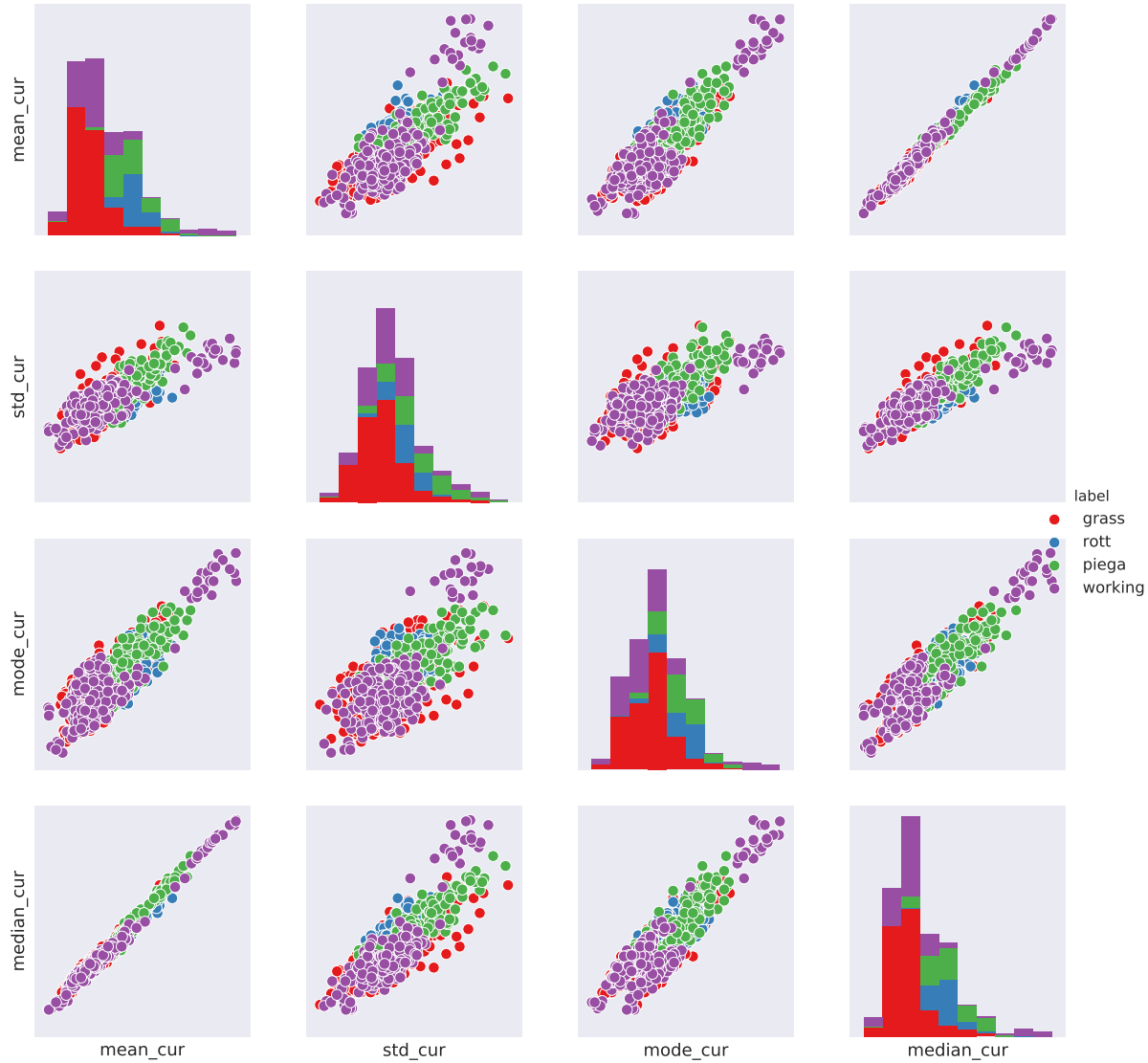
Scatterplot — Currents, Binary



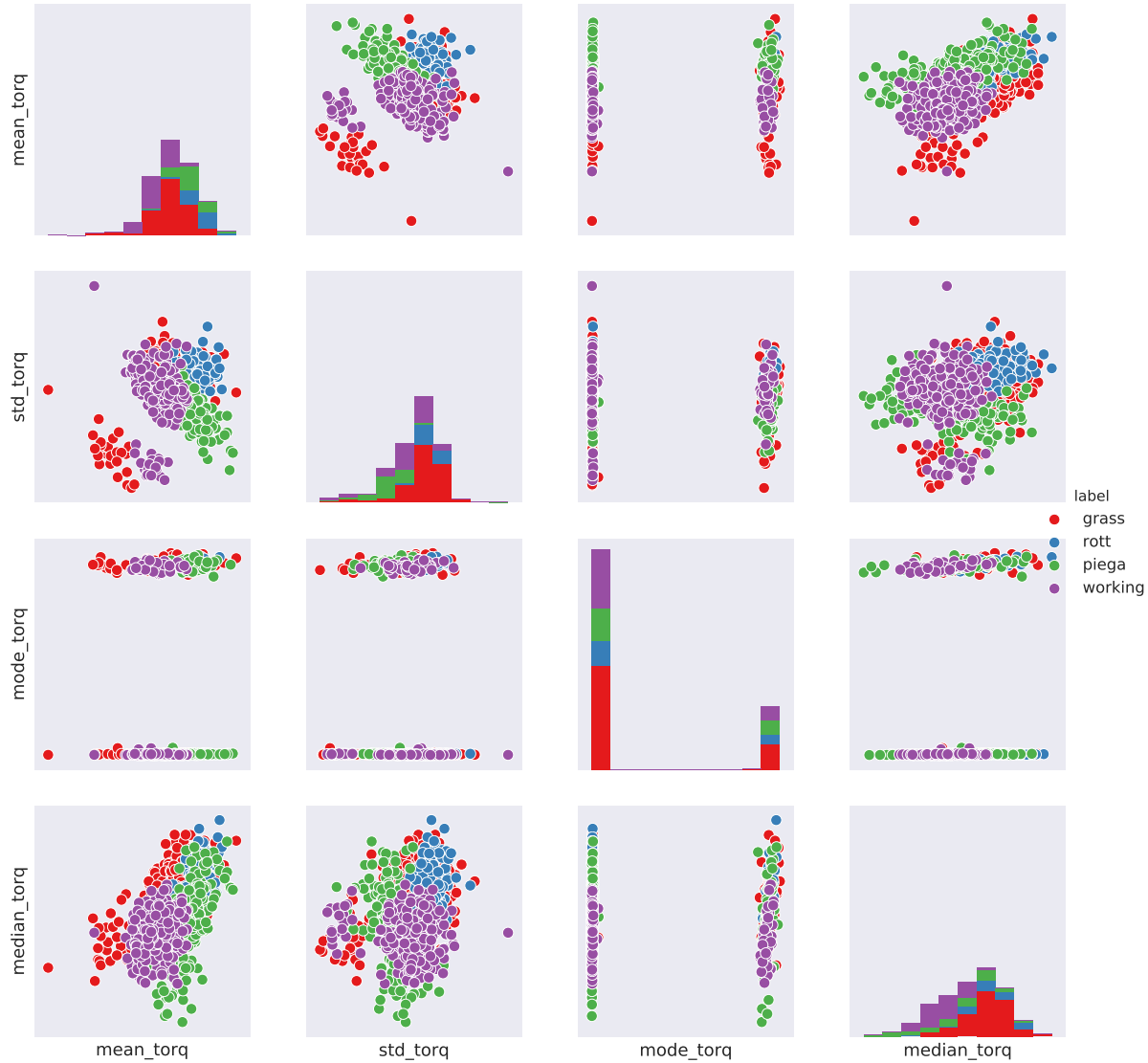
Scatterplot — Torques, Binary



Scatterplot — Currents, Multiclass



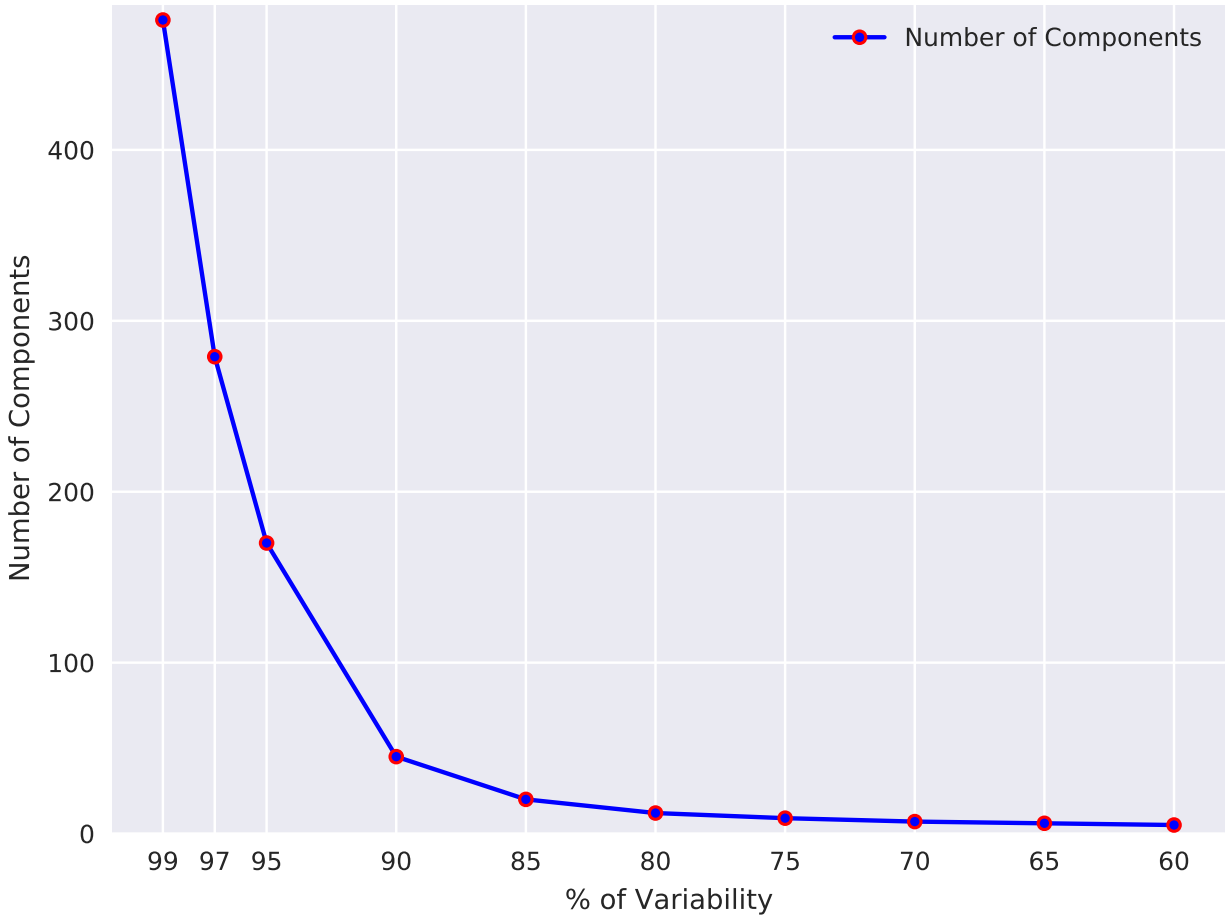
Scatterplot — Torques, Multiclass



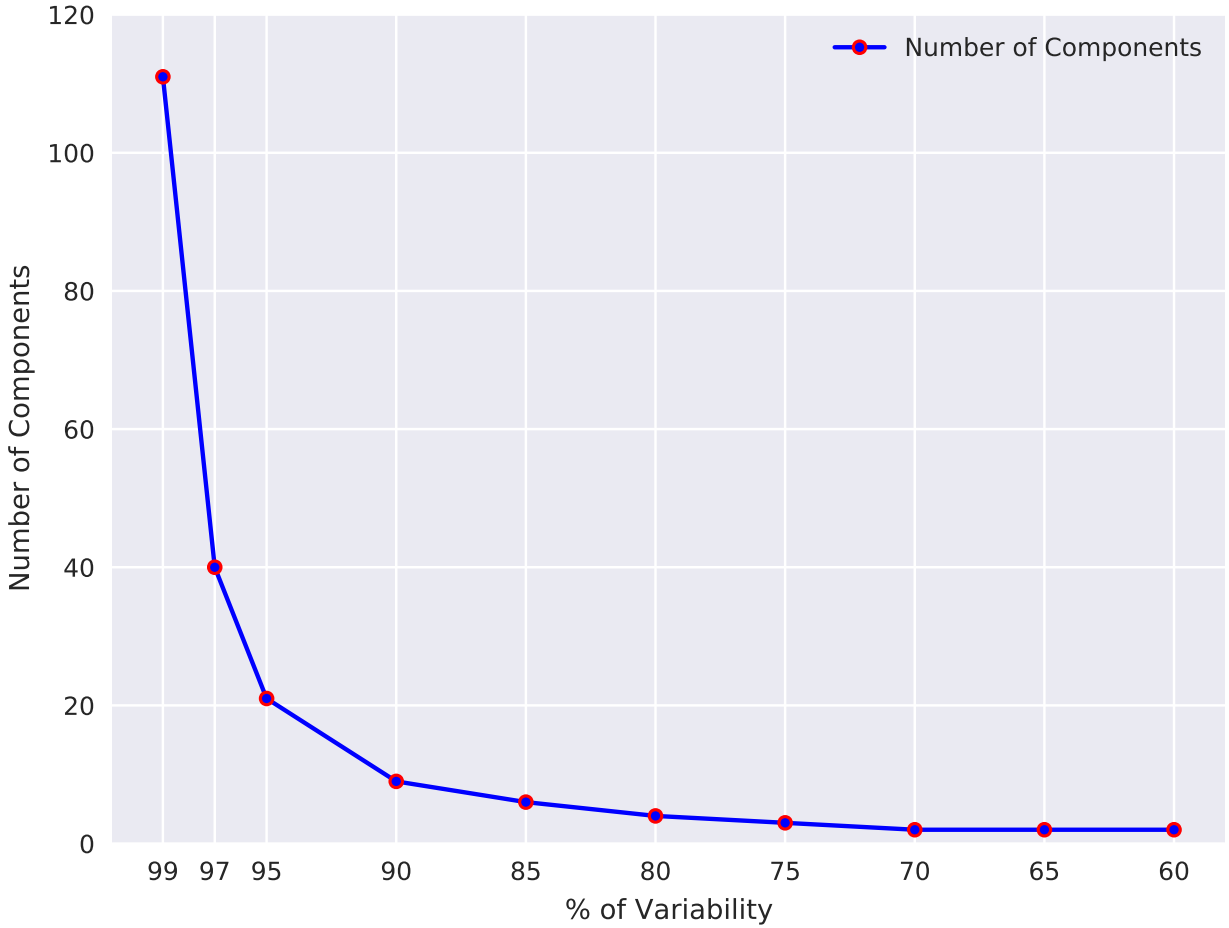
Exploratory Analysis: Second Approach

- Principal Component Analysis (time-series data, p=*1524*)
  Identify the measurement times where we have maximum variability —>
  to be used later for classification/prediction

Principal Component Analysis: Currents, *p=1524, N=924*



Principal Component Analysis: Torques, *p=1524, N=924*



Conclusions

- First Approach
  Further improvements can be made by adding different features and/or transforming the input data, the current feature set doesn't show a clear separation between faulty and working motors.
- Second Approach
  Most of the variability can be captured with a relatively low number of components (10/20 vs p=1524).