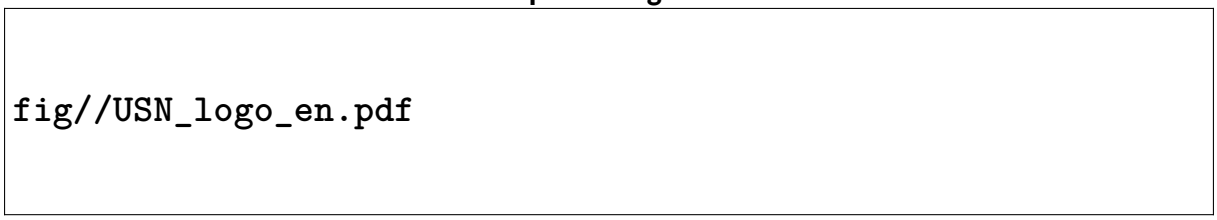
University of South-Eastern Norway

www.usn.no

IIA1420 2022

Assignment 2 - Data Visualization and Regression Models

<optional figure>



Lars Rikard Rådstoga

Faculty of Technology, Natural Sciences and Maritime Sciences

Campus Porsgrunn

Contents

1 Understanding the problem

The following chapter discusses and answers a few questions regarding the problem and the dataset, to better understand the following chapters.

1.1 What type of problem is it?

The problem at hand is in fact a regression task. The dataset available contains monitored sensory data, with noise, and can be considered labeled. The goal is to develop a model that can predict/extrapolate the RUL (Remaining Useful Life) of turbofan jet engines.

1.2 What category of machine learning is required?

As the dataset can be considered labeled, i.e. the columns in the dataset are labeled and have some physical meaning, the supervised learning category will be used.

1.3 What does each column in the dataset represent?

The dataset contains multiple columns, see figure 1.1 for brief descriptions of each column.

Table 1.1: Column descriptions

Engine	Identification number.
Cycle	Counted rotations since initialization.
Settings 1-3	How the systems configurations change over time.
Sensor 1-21	Various sensor measurements.
RUL	Remaining useful life.

1.4 Features and targets

Attributes are data types in the dataset that reflect the name of configured or measured values such as voltage set on a motor or pressure. Features are attributes bundled with a value. Targets are usually the information that is intended to predict. In this case features are all configuration and sensory data and the target is the RUL.

1.5 Not features or targets

Identification columns aren't features if they are all the same brand and make of fan. Time might be a feature.

2 Data Visualization

This chapter contains visualizations of the dataset that can be used to get better understanding of the system.

2.1 Histogram

All the available data was plotted, see figure 2.1.

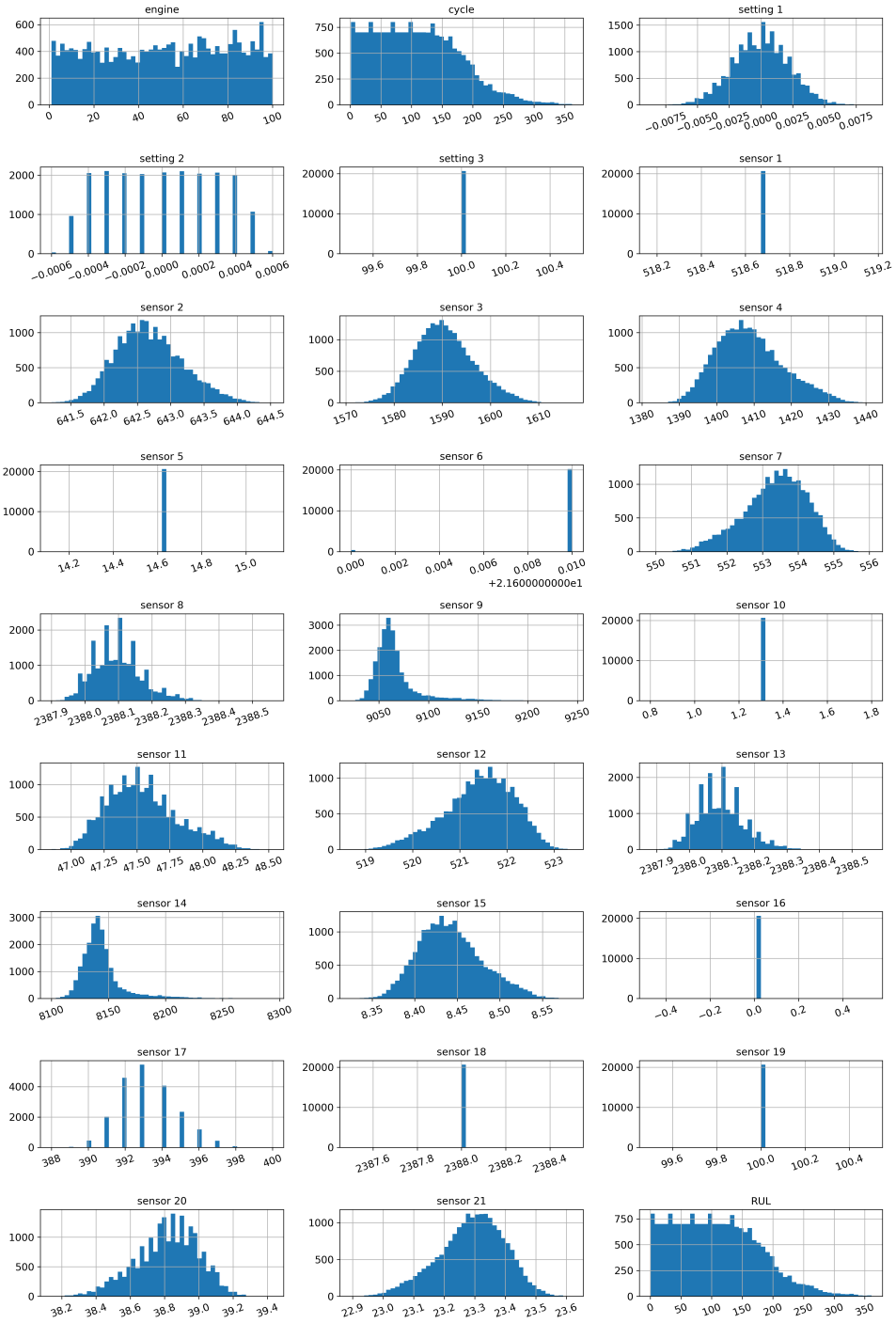


Figure 2.1: Histogram plot of all attributes.

3 Preprocessing and feature selection

This chapter discusses the dataset plotted in the previous chapter.*

3.1 Feature removal

Low correlation with feature: The engine (id) attribute is again not a feature or a target, and we can see that most engines have similar amounts of data points, engine attribute can safely be removed.

Highly correlated: Cycle and RUL are almost identical in terms of value frequency, which isn't strange since these are opposites, we should only use one of them.

High difference between feature values: Setting 2 and sensor 17 display gaps between recorded data. This may or may not be an issue.

Constant value:

Setting 3 and sensors 1, 5, 6, 10, 16, 18, 19, all have constant or close to constant values throughout the whole dataset. These should be removed.

Remaining attributes: Cycle or RUL, setting 1, 2, sensor 2, 3, 4, 7, 8, 9, 11, 12, 13, 14, 15, 17, 20, 21.

4 Regression models

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

5 Regression models with extended features

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

References

[1] *Chai - chat with ai*, https://chai.ml/, (Accessed on 09/14/2022).

[2] R. Murphy, R. Murphy and R. Arkin, *Introduction to AI Robotics* (A Bradford book). MIT Press, 2000, p. 4, ISBN: 9780262133838. [Online]. Available: https://books.google.ne/books?id=RVlnL%5C_X6FrwC.

[3] J. Sifakis, *Autonomous systems – an architectural characterization*, Nov. 2018.

[4] J. Černetič, S. Strmčnik and D. Brandt, 'Revisiting the social impact of automation,' *IFAC Proceedings Volumes*, vol. 35, no. 1, pp. 167–178, 2002, 15th IFAC World Congress, ISSN: 1474-6670. DOI: https://doi.org/10.3182/20020721-6-ES-1901.01648. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474667015400692.

[5] *Laveste arbeidsledighet på 14 år – e24*, https://e24.no/norsk-oekonomi/i/KzxVJ6/laveste-arbeidsledighet-paa-14-aar, (Accessed on 09/15/2022).

[6] *The unemployment rate fell to 3.5%, matching its lowest level in the last 50 years : Npr*, https://www.npr.org/2022/08/05/1116036160/the-unemployment-rate-fell-to-3-5-matching-its-lowest-level-in-the-last-50-years, (Accessed on 09/15/2022).

[7] *To av tre jordbærdyrkere på nes i ringsaker gir seg – nrk innlandet – lokale nyheter, tv og radio*, https://www.nrk.no/innlandet/to-av-tre-jordbaerdyrkere-pa-nes-i-ringsaker-gir-seg-1.15607791, (Accessed on 09/15/2022).

[8] J. Bryson and A. Winfield, 'Standardizing ethical design for artificial intelligence and autonomous systems,' *Computer*, vol. 50, no. 5, pp. 116–119, 2017. DOI: 10.1109/MC.2017.154.

[9] P. Gruhn, 'Human machine interface (hmi) design: The good, the bad, and the ugly (and what makes them so),' 2011.

[10] *Nozzle - wikipedia*, https://en.wikipedia.org/wiki/Nozzle, (Accessed on 09/19/2022).

[11] M. Endsley and E. Kiris, 'The out-of-the-loop performance problem and level of control in automation,' *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, Jun. 1995. DOI: 10.1518/001872095779064555.

Appendix A

Source code

The source code, Jupyter notebook, used to load, transform and plot data.