

Datasheet

Motivation

- For what purpose was the dataset created?

To build a model that predicts the likelihood of machine failure

- Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)? Who funded the creation of the dataset?

N/A. the dataset was taken from Kaggle.

Composition

- What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?

Six features and one output:

- Type
- Air temperature (K)
- Process temperature (K)
- Rotational speed (rpm)
- Torque (Nm)
- Tool wear (min)
- Machine failure (output)

- How many instances of each type are there? 10,000

- Is there any missing data? No

- Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor–patient confidentiality, data that includes the content of individuals' non-public communications)? No

Collection process

- How was the data acquired? Unknown to the authors of the datasheet

- If the data is a sample of a larger subset, what was the sampling strategy? Unknown to the authors of the datasheet

- Over what time frame was the data collected? Unknown to the authors of the datasheet

Preprocessing/cleaning/labelling

- Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section. Unknown to the authors of the datasheet

- Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? Unknown to the authors of the datasheet

Uses

- What other tasks could the dataset be used for? Only for predicting machine failure

- Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other risks or harms (e.g., legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms? **No**

- Are there tasks for which the dataset should not be used? If so, please provide a description. **No**

Distribution

- How has the dataset already been distributed? **It was available on Kaggle**

- Is it subject to any copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? **No**

Maintenance

- Who maintains the dataset? **Unknown to the authors of the datasheet**