

# Data Description Report

## Machine Predictive Maintenance Classification

---

### Data Quantity

#### ➤ File Format

- The given dataset is in CSV file format.
- A CSV file format is a delimited text file that separates values using commas.
- Each line in the file represents a data record, consisting of 10 fields separated by commas.

#### ➤ Data Gathering

Since real predictive maintenance datasets are generally difficult to obtain and difficult to publish, we present and provide a synthetic dataset that reflects real predictive maintenance encountered in industry to the best of our knowledge.

We can find source of data from given link :

<https://drive.google.com/file/d/1xFhSYUi938NJaonqXn3mh5jagn4sCJ1o/>

#### ➤ Data Dimension

Fields	10
Records	10,000

## Data Quality

### ➤ Characteristics Of Data

The dataset contains features that are relevant to the business question of predictive maintenance for industrial machines. The features include **air temperature, process temperature, rotational speed, torque, and tool wear**, which are all important indicators of potential machine failure.

### ➤ Data Description & Types

- UID: A unique identifier ranging from 1 to 10,000.
- productId: A string consisting of a letter L, M, or H for low (50% of all products), medium (30%), and high (20%) as product quality variants and a variant-specific serial number.
- air temperature [K]: Generated using a random walk process later normalized to a standard deviation of 2 K around 300 K.
- process temperature [K]: Generated using a random walk process normalized to a standard deviation of 1 K, added to the air temperature plus 10 K.
- rotational speed [rpm]: Calculated from power of 2860 W, overlaid with a normally distributed noise.
- torque [Nm]: Torque values are normally distributed around 40 Nm with an  $\sigma = 10$  Nm and no negative values.
- tool wear [min]: The quality variants H/M/L add 5/3/2 minutes of tool wear to the used tool in the process.
- Failure or Not: A binary label that indicates whether the machine has failed in this particular data point for any of the following failure modes are true.
- Type of Failure: A categorical label that indicates the type of failure, if any.

This dataset contains the **numeric and strings** values.

Find in column Field							
<input type="checkbox"/>	Field	Measure ⓘ		Role ⓘ		Value mode ⓘ	Values
<input type="checkbox"/>	# UDI	Continuous	▼	Input	▼	Instantiated	▼ 1, 10000
<input type="checkbox"/>	abc Product ID	Typeless	▼	None	▼	Instantiated	▼
<input type="checkbox"/>	abc Type	Nominal	▼	Input	▼	Instantiated	▼ H, L, M
<input type="checkbox"/>	# Air temperature [K]	Continuous	▼	Input	▼	Instantiated	▼ 295.3, 304.5
<input type="checkbox"/>	# Process temperature [K]	Continuous	▼	Input	▼	Instantiated	▼ 305.7, 313.8
<input type="checkbox"/>	# Rotational speed [rpm]	Continuous	▼	Input	▼	Instantiated	▼ 1168, 2886
<input type="checkbox"/>	# Torque [Nm]	Continuous	▼	Input	▼	Instantiated	▼ 3.8, 76.6
<input type="checkbox"/>	# Tool wear [min]	Continuous	▼	Input	▼	Instantiated	▼ 0, 253
<input type="checkbox"/>	# Target	Continuous	▼	Input	▼	Instantiated	▼ 0, 1
<input type="checkbox"/>	abc Failure Type	Nominal	▼	Input	▼	Instantiated	▼ Heat Dissipation Fail...

SPSS Modeler: Type Node

## ➤ Data Statistics

Name	Measurement	Type	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
UDI	Continuous	Long	1	10,000	5000.5	2886.896	0	0	10000
Product ID	Nominal	String	--	--	--	--	--	255	10000
Type	Nominal	String	--	--	--	--	--	3	10000
Air temperature	Continuous	Double	295.30	304.50	300.005	2.000	0.114	0	10000
Process temperature	Continuous	Double	305.70	313.80	310.006	1.484	0.015	0	10000
Rotational speed	Continuous	Long	1,168	2,886	1538.77	179.284	1.993	0	10000
Torque	Continuous	Double	3.800	76.600	39.987	9.969	-0.010	0	10000
Tool wear	Continuous	Long	0	253	107.951	63.654	0.027	0	10000
Target	Continuous	Long	0	1	0.034	0.181	5.150	0	10000
Failure Type	Nominal	String	--	--	--	--	--	6	10000

## ➤ Key Attributes / Priority Fields

It could be difficult to prioritize field without any insight knowledge of machine. But we can analyse the below fields for our projects:

- Air temperature
- Process temperature
- Rotational Speed
- Torque
- Tool wear

## ➤ Non-Priority Fields

There are **UID** and **Product ID** attributes which would not helping to analyse the feature of machine and patterns. It is only a information to specific record to identify unique transaction.