BEN GURION UNIVERSITY OF THE NEGEV

Faculty Of Engineering Sciences

Predictive Maintenance Data Visualization

By:

Yogev Attias – 316614254

Shir Michal Grief – 318438124

GitHub Repo

https://github.com/yogevat/Data_Visualization_Project

April 24

BEN GURION UNIVERSITY OF THE NEGEV

Faculty Of Engineering Sciences

Table of contents

Table of Figures

April 24

1. Introduction

Predictive maintenance (PdM), also known as Condition Based Maintenance (CBM) is a proactive maintenance strategy used in various industries to predict when machinery or equipment is likely to fail. Predictive maintenance is important for industry because it helps companies avoid costly downtime, reduce maintenance costs, and improve operational efficiency. This approach allows organizations to schedule maintenance activities only when necessary, minimizing the impact on production and maximizing equipment uptime. This project, merging advanced data analytics with traditional manufacturing, aims to elevate operational excellence. Utilizing a detailed dataset from an industrial environment, our objective is to create a predictive maintenance framework that anticipates equipment failures and inefficiencies. In the context of predictive maintenance, data visualization plays a crucial role in helping organizations understand the health of their equipment and make informed decisions. By visualizing key companies can quickly identify patterns, anomalies, and trends that may indicate potential issues with their equipment.

2. Understanding Data

In the realm of predictive maintenance, the key to unlocking the full potential of industrial processes lies in a deep understanding of the data. Each entry in our dataset is a snapshot of a dynamic process. From air and process temperatures to rotational speeds and torque, each variable encapsulates a story of the machine's operation. Understanding this data involves deciphering these processes, discerning patterns and anomalies that might otherwise go unnoticed. This requires a meticulous examination of the relationships between different parameters, an exploration of how they interact and influence one another. The challenge is to transform these raw numbers into insights that can predict when a machine might fail. By mastering the language of this data, we empower ourselves to anticipate problems, optimize performance, and drive innovation in industrial processes.

3. Understanding Business

Understanding the business context is pivotal in ensuring that our predictive maintenance project transcends beyond theoretical models to practical, impactful solutions. In the fast-paced industrial sector, operational efficiency and equipment reliability are not just operational goals, but they are also critical drivers of financial performance and competitive advantage. Every machine failure avoided, every hour of downtime reduced, translates into tangible business benefits - be it in the form of cost savings, improved productivity, or enhanced product quality. Therefore, our approach must be rooted in a deep comprehension of the industry's dynamics, challenges, and objectives. We need to align our predictive insights with the strategic goals of the business, ensuring that our recommendations resonate with the needs of decision-makers and stakeholders. This business-centric perspective is crucial in transforming our data-driven insights into actionable strategies that can propel the business forward, delivering value that is both measurable and meaningful.

4. Data Explanation

4.1 Identifier Data:

- UDI (Unique Identifier): A unique number assigned to each record, ensuring each entry in the dataset is distinct and can be referenced individually.

- Product ID: A unique identifier for each product.

4.2 Categorical Data:

- Type: This represents different categories or types of products. (L- Low, M-Medium, H-High)

4.3 Continuous Numerical Data:

- Air temperature [K]: This is a continuous variable representing the air temperature in Kelvin during manufacturing.

- Process temperature [K]: This is a continuous variable representing the temperature of the manufacturing process itself.

- Rotational speed [rpm]: Measures the speed of equipment in rotations per minute.

- Torque [Nm]: This is the force exerted by the machinery, measured in Newton-meters.

- Tool wear [min]: Indicates how long a tool has been used, in minutes.

4.4 Binary Data:

- Machine failure: This is a binary indicator (0 or 1), where 1 indicates a machine failure and 0 indicates normal operation. The possible failures are:

TWF (Tool Wear Failure), HDF (Heat Dissipation Failure), PWF (Power Failure), OSF (Overstrain Failure), RNF (Random Failures).

These are binary indicators (0 or 1) representing different types of machine failures or issues. Each one specifies a particular kind of failure.

5. Visualization

Visualizing this data helps us uncover patterns and correlations that are essential for understanding machine performance and reliability.

5.1 Correlations of Industrial Machine Parameters with Failure Events

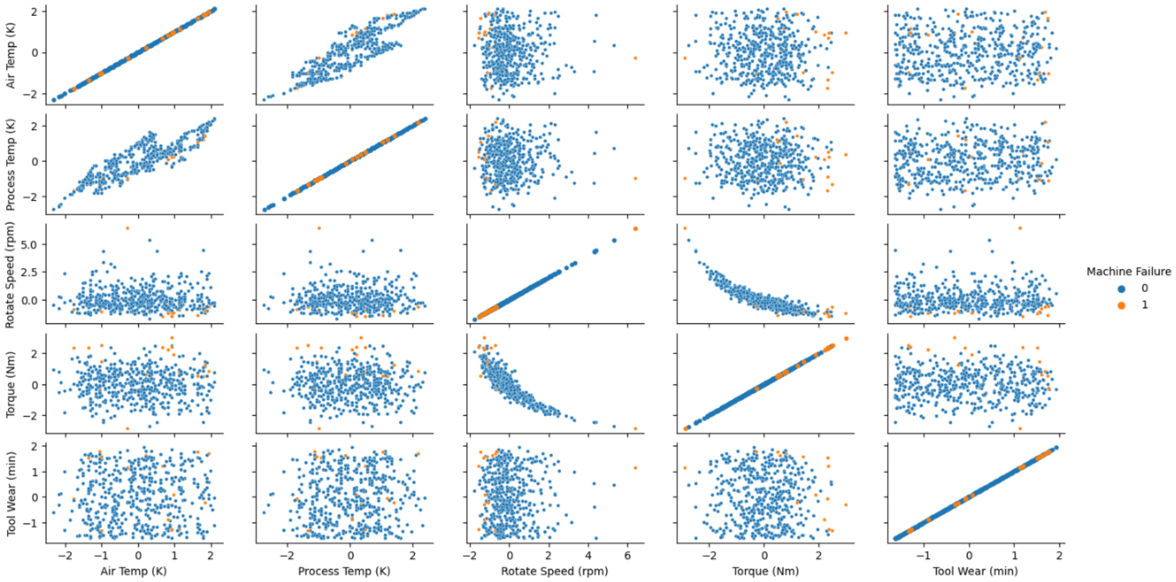

Figure 1 Correlations of Industrial Machine Parameters with Failure Events

Idiom	Scatter Plot Matrix (SPLOM)
What: Data. (continuous variables)	Table.
Why: Analyze – Produce	Find correlations, trends, outliers.
How: Encode	Scatter plots in 2D matrix alignment with colors.

Conclusion: As we can see from the figure there is a positive linear correlation seems to be evident between Air Temperature and Process Temperature, as well a negative relationship between Rotational Speed and Torque.

5.2 Density Distributions of Machine Parameters by Type

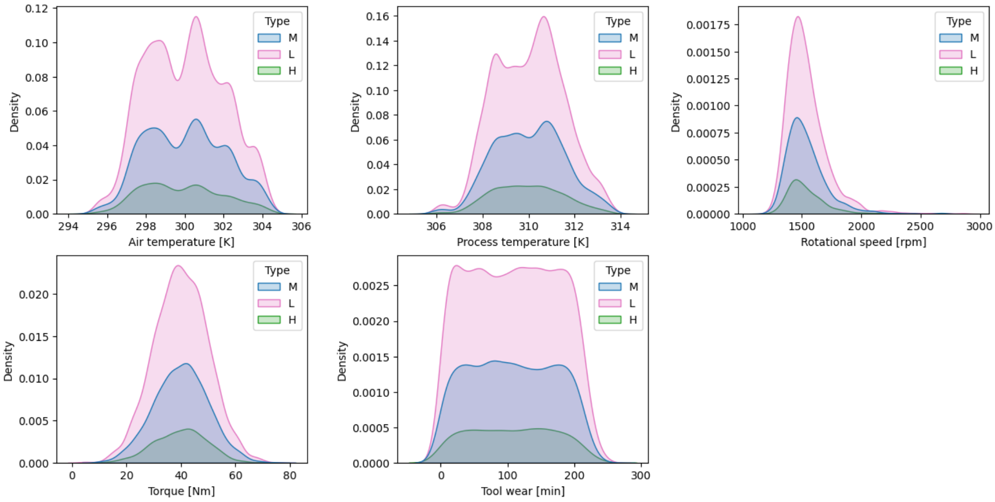

Figure 2 Density Distributions of Machine Parameters by Type

Idiom	**Kernel Density Estimate (KDE)**
What: Data. (continuous variables)	Distribution of single variable by showing different types
Why: Analyze	Comparing the distributions of machine parameters between the different types
How: Encode	Each variable density is encoded using areas under the curve, with color indicating the type or condition

Conclusion: This suggests that 'H' type conditions or operations are the most strenuous or demanding on the equipment, which could potentially be associated with a higher risk of failure or the need for more frequent maintenance due to the higher tool wear and operating temperatures.

5.3 Interactive Analysis of Process Temperature Variability by Machine Failure Type

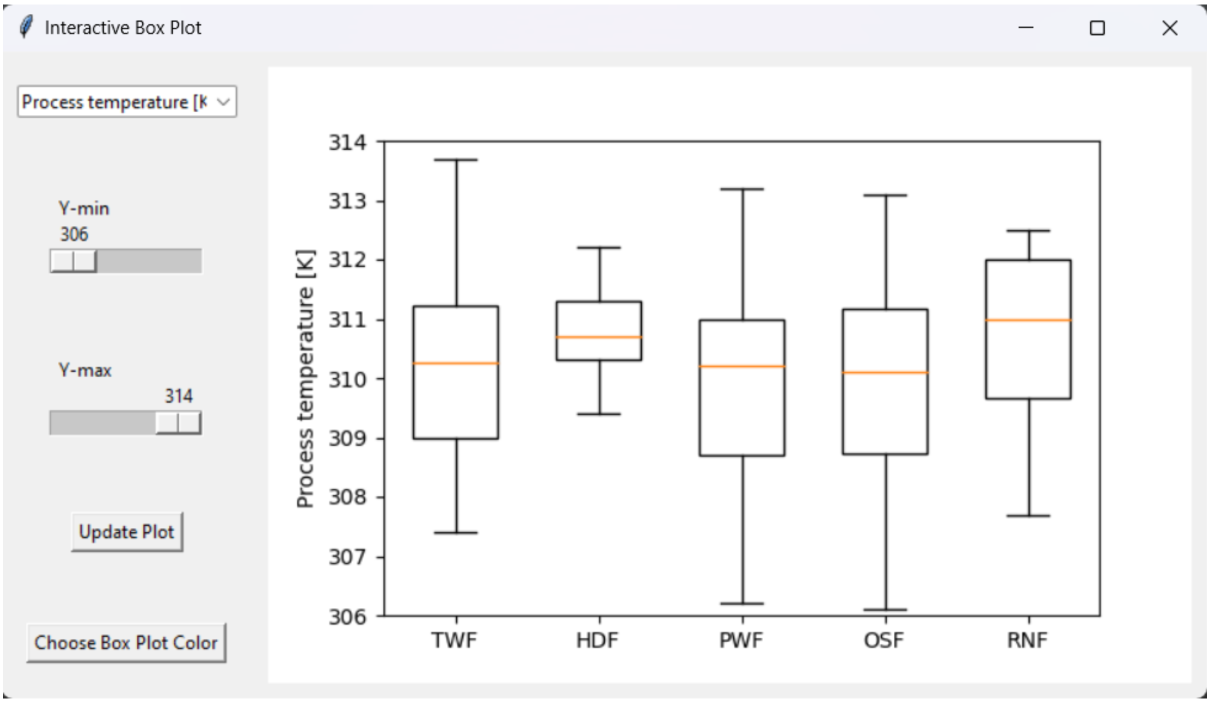

Figure 3 Interactive Analysis of Process Temperature Variability by Machine Failure Type

* This interface constitutes a graphical user interface (GUI) featuring a lateral slide bar, integrated on the left segment of the display. The user is empowered to select a variable of interest via this slide bar. Correspondingly, the displayed range is dynamically adjustable through a sliding mechanism, with the default settings configured to represent the minimum and maximum values of the selected variable. Furthermore, the interface offers customization options, enabling users to modify the chromatic scheme of the box plot representation (slide bar script).

Idiom	**Interactive Box plot**
What: Data. (continuous and Categorial variables)	Each box plot represents a statistical summary of the continuous data for each category.
Why: Analyze	Comparing the median, spread, and outliers of process temperature across the different categories.
How: Encode, Interaction	The central tendency of the data is encoded with a line inside each box, the spread is encoded with the box itself, and outliers are encoded with points.

Conclusion: There is no category with a median that is exceptionally high or low compared to the others, which could suggest that process temperature is not a distinguishing factor between these categories, or it is controlled within a narrow range.

5.4 Interactive Process Temperature vs. Power Output Colored by Error Type

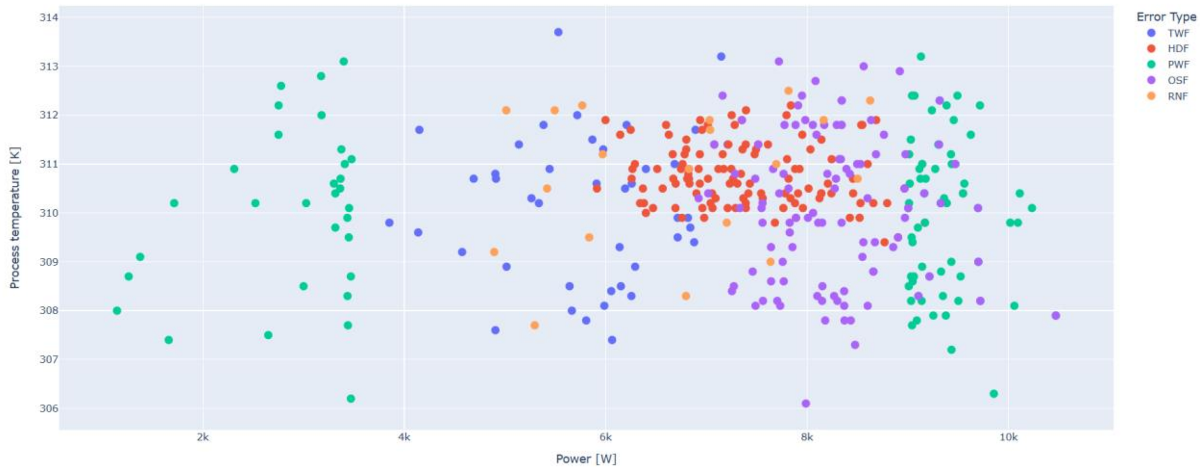

Figure 4 Process Temperature vs. Power Output Colored by Error Type

* This is an interactive plot designed for user engagement, whereupon selecting any given type of failure, its corresponding data visualization is immediately rendered visible (Classifier script).

Idiom	**Interactive Scatter**
What: Data. (continuous and Categorial variables)	Two-dimensional scatter plot, and dimension of color.
Why: Analyze	Users can identify clusters or outliers within the data that may suggest specific conditions or anomalies related to error types
How: Encode, Interaction, Manipulate	The quantitative variables are encoded using positional axes. The error type is encoded using color. Manipulating: calculating the power by using the formula of Torque twice Rotation Speed

Conclusion: Power and process temperature appear to be related, and the distribution of different error types across these variables suggests that each error type has a distinct pattern with respect to power and temperature conditions. Understanding these patterns could be valuable for predictive maintenance or for optimizing operational parameters to minimize the occurrence of errors.

5.5 Machine Type Frequency and Failure Incidence

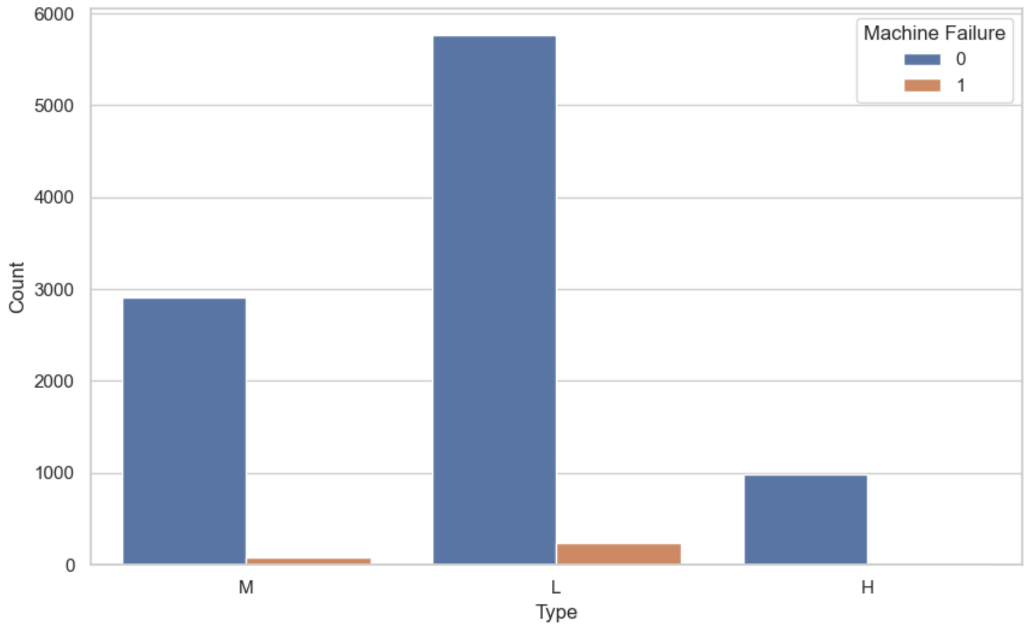

Figure 5 Machine Type Frequency and Failure Incidence Machine Type Frequency and Failure Incidence

Idiom	**Interactive Scatter**
What: Data	Table: one quantitative value attribute, Two categorical key.
Why: Task	Lookup and compare values.
How: Encode, Manipulate	The bar height encodes the count of machines, with color encoding the failure status (e.g., blue for no failure and orange for failure). Manipulating the probability by calculating the mean (as a percentage) of a 'Target' column in the Data Frame.

Conclusion: Type H has the highest apparent rate of failure, Type L has the highest number of operations with the lowest relative failure rate, and Type M is intermediate in both total counts and failure rates. This information could be crucial for operational planning, maintenance scheduling, and risk management.

5.6 Probability of Machine Failure Across Various Operational Parameters

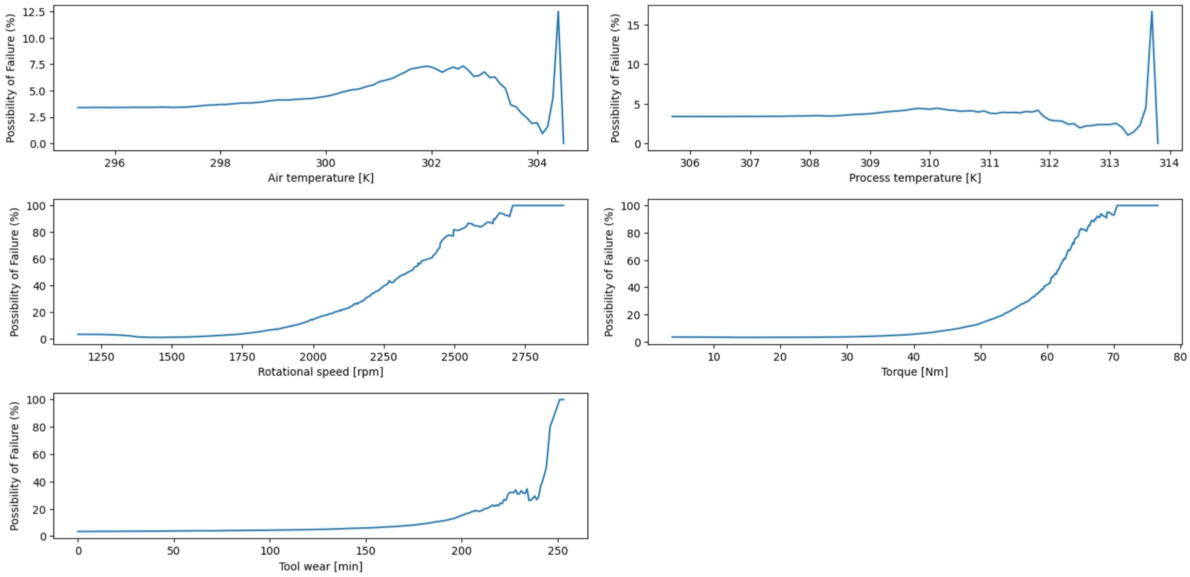

Figure 6 Probability of Machine Failure Across Various Operational Parameters

Idiom	**Line**
What: Data	Probability as function of continuous variable
Why: Analyze	Identifying specific points where there are significant changes or anomalies in the parameter values, such as spikes in tool wear or drops in temperature.
How: Encode	The line graph uses position on a two-dimensional plane to encode the value of the .machine parameters

Conclusion: For air and process temperatures, the probability of machine failure is not uniform across the range; instead, it is prone to spike at values. The probability of failure increases with higher rotational speeds, particularly beyond a specific speed. As torque increases, the likelihood of failure also increases, and this relationship appears to be more linear. There is a significant increase in failure probability at higher levels of tool wear, indicating a threshold effect where the risk of failure escalates sharply. These relationships suggest that careful monitoring and control of these parameters could be crucial in preventing machine failure. Maintaining operational parameters within safe ranges could potentially reduce the risk of failure. It also suggests that preventive maintenance could be scheduled around the critical points identified in the tool wear and rotational speed graphs to avoid sudden increases in the probability of failure.