Engineering Insights

1. Data Stream System

While building the streaming dashboard, we realized that a real-time monitoring system is very different from normal data analysis in a notebook. In offline analysis, we can load the whole dataset and compute results at once. In contrast, a data stream system only receives one data point at a time and must immediately decide how to react.

Because of this limitation, the system needed mechanisms such as sliding windows, minimum work points, and an alert cooldown. At the beginning, the system triggered alerts continuously for the same abnormal behavior. Adding a cooldown period allowed the system to “remember” previous alerts and prevented repeated notifications. This showed that a streaming monitoring system must maintain state information over time instead of only processing individual data points.

Another important lesson was that data flow stability was more important than model accuracy. If the database connection lagged or the data arrived out of order, the alert logic behaved incorrectly even when the model was correct. Therefore, the project demonstrated that real-time machine learning applications depend heavily on reliable data pipelines, not only on algorithms.

The live visualization also became an essential part of the system. The plot helped verify whether an alert was reasonable and allowed us to understand machine behavior in context. Without the visualization, the alert messages alone were difficult to interpret. This means the dashboard is part of the decision-support system rather than just a display.

---

2. Linear Regression

During this project, linear regression was not mainly used to predict exact values but to describe normal machine behavior. A simple regression line allowed us to observe the trend of force measurements during working periods. We found that the slope of the regression line was more informative than the prediction error. When the slope gradually increased, it often indicated rising mechanical resistance or potential wear.

We also discovered that model performance depended strongly on data preprocessing. When idle periods and noisy segments were included, the regression parameters became unstable. After filtering non-working data and correctly segmenting work intervals, the results improved significantly even without changing the algorithm. This showed that good data preparation is more important than choosing a complex model.

Finally, regression alone could not reliably detect anomalies. Some abnormal events produced large force values but did not strongly affect the regression trend, while gradual degradation affected the trend but not individual peaks. Combining regression trend analysis with statistical thresholds provided a more stable detection method and reduced false alarms.

Overall, the regression model worked best as an interpretable indicator of machine condition rather than a pure prediction tool. It supported maintenance decisions by showing how the machine behavior changed over time.