

# Azure Predictive Maintenance

Lesrene, Kelsey, Toluwani

DS4420 Final Project

## Abstract

Predictive maintenance has become an important focus for companies that want to prevent unexpected breakdowns and reduce the cost of machine downtime. Instead of waiting for a component to fail or replacing parts on a fixed schedule, our goal is to use historical data to predict which component on a machine is most likely to fail next.

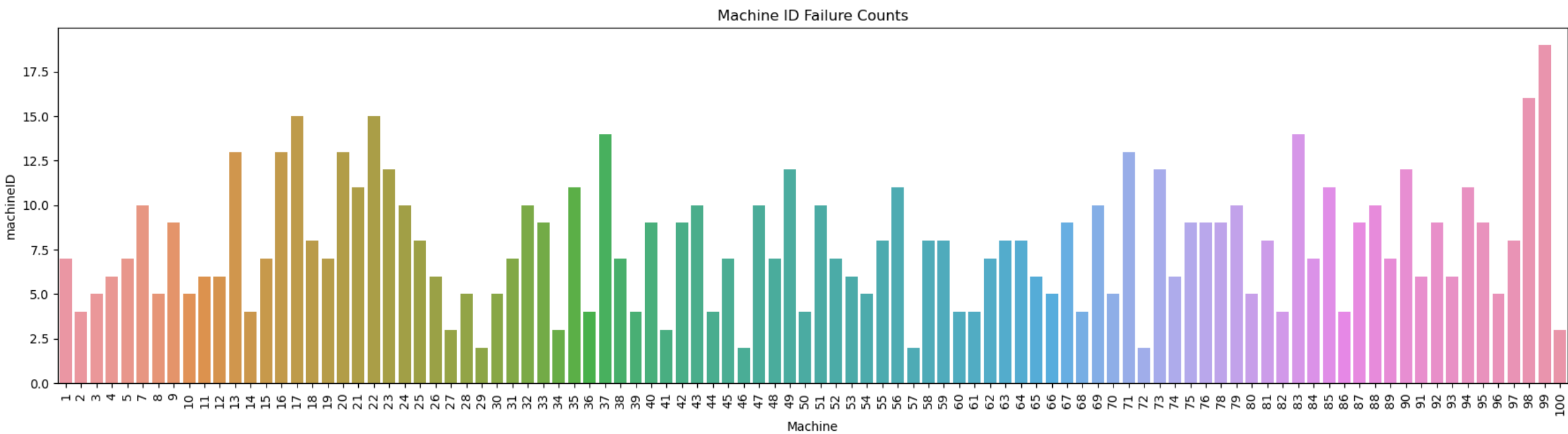
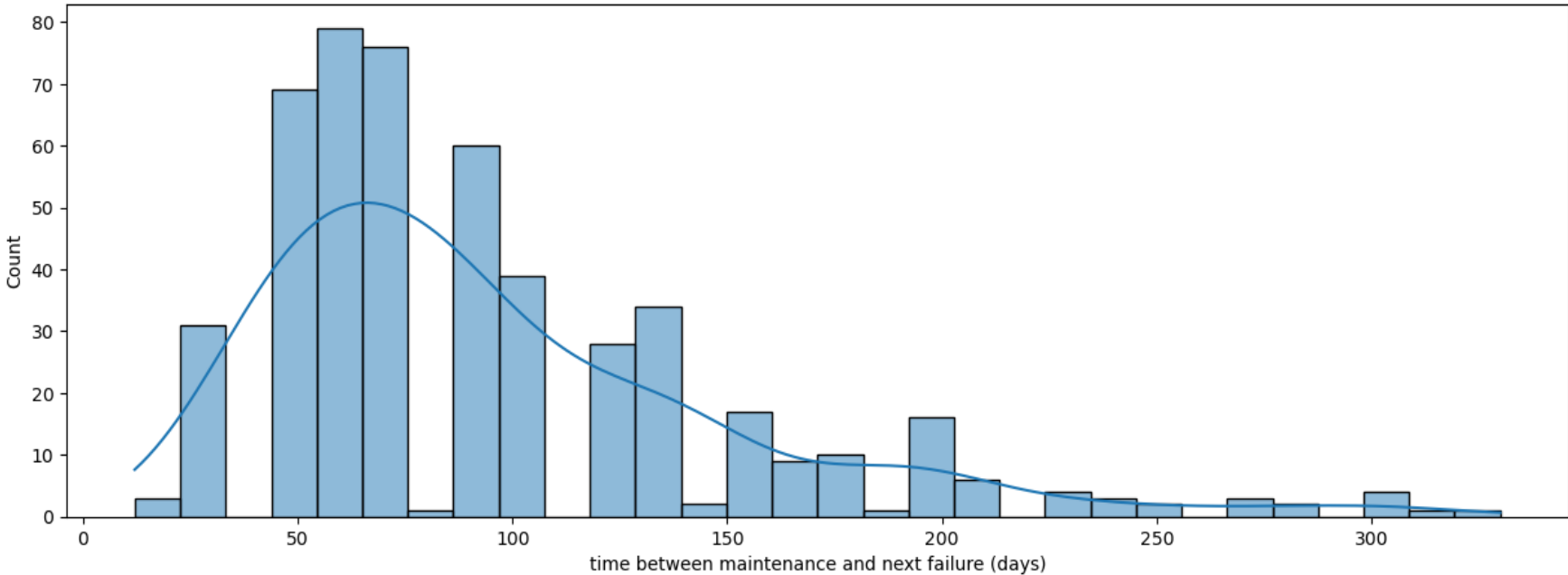
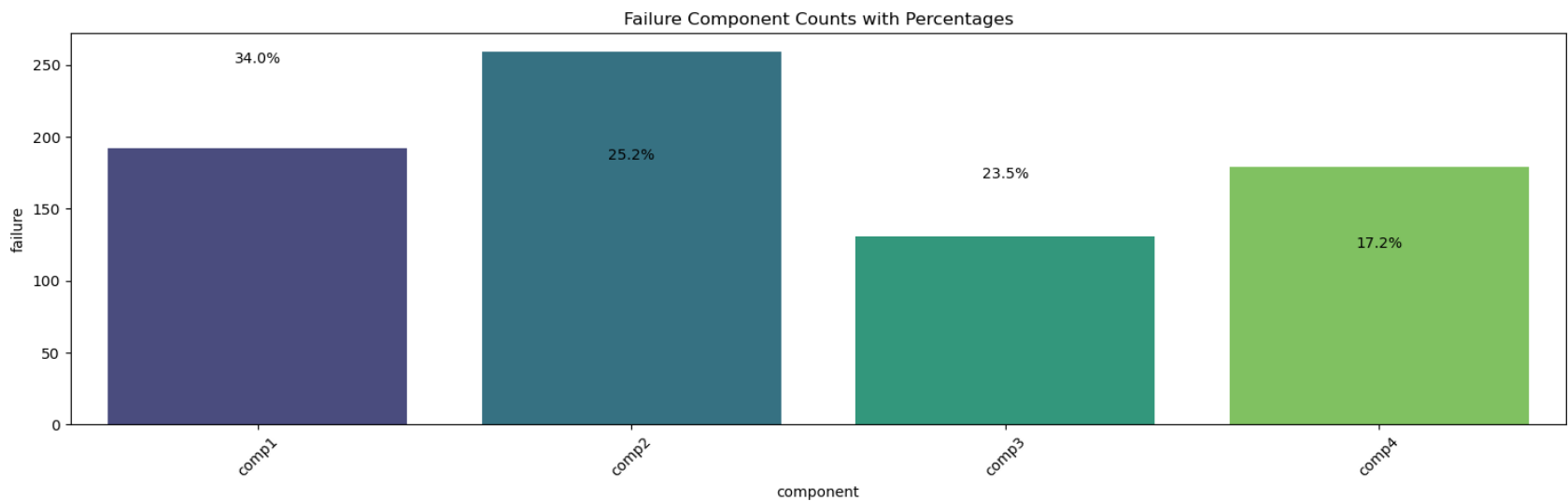
In this project, we combine two modeling approaches that each provide different strengths. The first uses collaborative filtering, where machines are treated like “users” and components are treated like “items.” The idea is that machines with similar failure histories can help predict future failures even when the data is sparse. The second approach uses Bayesian pattern discovery to identify sequences of errors and events that tend to appear shortly before failures. Together, these methods create a more complete and interpretable view of failure risk.

## Motivation

Industrial machines produce a large amount of data from sensors to error logs. When this information is used effectively, it can help companies avoid unplanned downtime and schedule repairs more efficiently. The challenge we focus on is determining which component on each machine is most likely to fail based on patterns seen across the entire fleet.

We work with a dataset that includes error events, component failures, telemetry readings, maintenance history, and machine attributes such as age and model type. Our exploration of the data shows several important patterns. Some components fail much more frequently than others. Some machines experience repeated failures while others fail rarely. Older machines tend to show higher failure risk. There are also clear differences between model types, and the time between maintenance and the next failure varies widely but often clusters between forty and eighty days.

These patterns helped guide our modeling choices and provided evidence that both collaborative filtering and Bayesian methods could be effective on this type of industrial data.



## Methodology

### Collaborative Filtering Approach

We treat each machine as if it were a “user” and each component as an “item.” Whenever a component fails on a given machine, that interaction becomes part of a machine component matrix. This structure allows us to apply collaborative filtering, where the model learns to identify machines that behave similarly based on the pattern of their failures. Machines that share similar patterns produce useful information even when their data is sparse.

We build representations that summarize how machines relate to each other, taking into account failure frequency and similarities in component behavior. Once the model learns these patterns, it can generate a ranking of which components appear most likely to fail next for any specific machine.

### Bayesian Pattern Discovery

In parallel with collaborative filtering, we developed a survival model using brms. We combined failures and maintenance timestamps into a single timeline for each machine and component. Failures were treated as events, while maintenance acted as censored observations that reset the clock. For every interval between maintenance and the next event, we calculated the time-to-event in hours.

We then merged these intervals with several sets of engineered features. Error logs were converted into counts per machine and error type. Telemetry data were aggregated within each interval to compute average voltage, rotation, pressure, and vibration. We also added component-level features such as how many times the component had been maintained and how long it had been since the last maintenance.

We used a Cox survival model in R (via the brms package) to understand how these features affect the risk of failure. The model learned which conditions increase or decrease the chance of a part failing soon.

Then we used the model to predict failure risk for each part of a specific machine. The output gives a risk score that helps rank the parts from most risky to least risky.

## Results and Evaluation

Our data exploration showed clear patterns. Some parts failed a lot more often than others. Some machines failed many times while others hardly failed at all. Older machines tended to break more often. Time between repairs and failures was not random and usually fell into a consistent range.

The collaborative filtering model was good at finding groups of machines that had similar failure patterns. It also highlighted a small set of parts that are usually high-risk.

The survival model explained why certain parts are risky. For example, long gaps between maintenance and higher pressure or vibration readings increased the chance of failure. When we ranked the parts for a single machine, we got a clear order showing which parts were more likely to fail soon.

Together, both models gave useful and understandable results.

## Next Steps

The two modeling approaches contribute different strengths. Collaborative filtering naturally handles sparse data and discovers patterns that only appear when looking across many machines. The Bayesian survival model focuses on the timing of events and provides uncertainty estimates that are useful for maintenance planning. Using these methods together improves both the interpretability and accuracy of predictions.

Several future steps could strengthen this work. A natural extension is to incorporate deep learning models that directly process telemetry time-series instead of aggregated values. Another direction is to develop a hybrid model that uses Bayesian risk estimates as features inside the collaborative filtering system. Adding richer telemetry streams such as vibration spectra or temperature trends may also improve early detection. Finally, these predictions could be integrated into a dashboard that continuously updates risk scores and helps maintenance teams schedule interventions more effectively.