Machine Learning and Data Mining

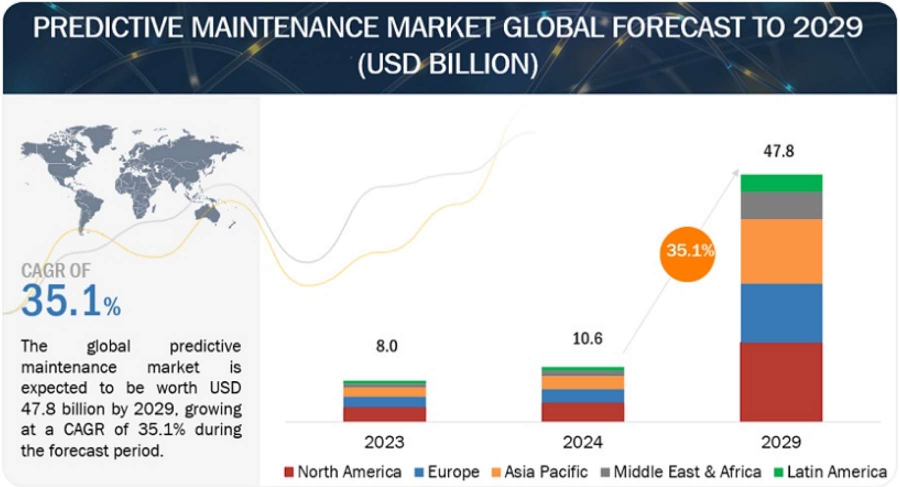
CRISP Case Study

Claudio Sartori

DISI
Department of Computer Science and Engineering – University of Bologna, Italy
claudio.sartori@unibo.it

Caveat

This presentation uses many terms of machine learning. Those terms may be obscure at this moment, they will be explained during the entire course.



PREDICTIVE MAINTENANCE MARKET GLOBAL FORECAST TO 2029 (USD BILLION)

CAGR OF **35.1%**

The global predictive maintenance market is expected to be worth USD 47.8 billion by 2029, growing at a CAGR of 35.1% during the forecast period.

Source https://www.marketsandmarkets.com/

Outline

1. Business Understanding 4
   - Informal discussion 5
   - Insights from the informal discussion 12
2. Data understanding 19
   - Example data 23
3. Data Preparation 32
4. Modelling 41
5. Evaluation 49
6. Deployment 58
7. Final Notes 67
8. 71

Production Managers on CNC Operations I

Manager 1

Well, our CNC machines are basically the heart of our production. They're the automated tools that actually make the precision parts - all the critical stuff that goes into our assemblies. We use them across the board - for aerospace-type components, automotive jobs, even some electronics housings and heavier industrial pieces. They're supposed to run nonstop, 24/7 if possible, because once they stop, we start losing time and money right away.

Manager 2

Exactly. These machines are built for high precision, so they can't really afford any inconsistency. A few microns off and the whole batch might be scrap. But the tricky part is they need to keep that level of accuracy while running continuously, which puts a lot of strain on the mechanical parts - spindles, bearings, motors, the works. Over time, they just wear out, no matter how careful we are.

Production Managers on CNC Operations II

Manager 1

Yeah, and the environment doesn't help much either - temperature shifts, vibrations, maybe even humidity - all of that makes the wear worse. Sometimes we see degradation sooner than expected, and it's not always clear if it's the part, the setup, or the operator. Unplanned downtime hits us hard - missed deadlines, production delays, extra costs for repairs. It's frustrating because these machines are meant to be reliable, but when one goes down, the whole line feels it.

Manager 2

Right, they're both robust and fragile at the same time - built for precision but sensitive to small things. That's the contradiction of running CNC operations: they're high-tech and reliable, until they're not.

Managers Discuss Maintenance Challenges I

Manager 1

You know, the biggest headache with these CNC machines is that they just fail out of the blue. One minute everything's running fine, and the next, a spindle jams or a drive motor trips an alarm, and we're suddenly behind schedule. We try to keep up with the regular maintenance plan, but even with all the checklists and service intervals, things still break down when you least expect it.

Manager 2

Yeah, the old approach - either waiting until something fails or sticking strictly to the schedule - doesn't really cut it anymore. Sometimes we replace parts that still have plenty of life left, and sometimes something goes bad in between planned maintenance. It's hit or miss, and both ways cost us money.

Managers Discuss Maintenance Challenges II

Manager 1

And speaking of money, those unplanned stoppages aren't cheap. When a machine goes down unexpectedly, it's not just the repair - we're talking about expensive components, lost production hours, idle operators, and even delayed deliveries. It adds up fast. You lose efficiency, and customers start asking questions about deadlines.

Managers Reflect on Data and Predictive Maintenance I

Manager 2

Exactly. And the irony is, these machines collect tons of data - every temperature, vibration, load reading, you name it. The sensors are constantly recording everything, but making sense of all that data? that's another story. We'd need proper analytics or data mining tools to get anything useful out of it.

Manager 1

Yeah, it's like drowning in information but starving for insight. We know there's something in that data that could help us predict failures better, but right now it's too complex to handle manually.

Managers Reflect on Data and Predictive Maintenance II

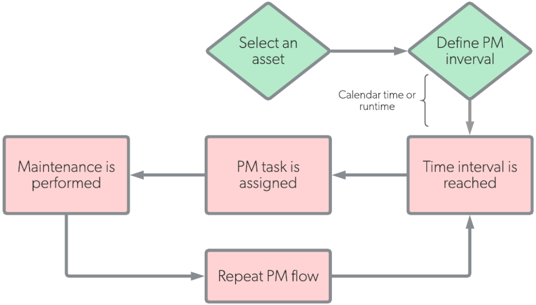
Manager 2

And even if we had better predictions, scheduling maintenance is still tricky. You want to keep the machines running as much as possible, but if you push too hard, you risk breakdowns. Do too much maintenance, and you're wasting resources and downtime. It's a balancing act - one we're constantly trying to get right, but it feels like we're always reacting instead of planning ahead.

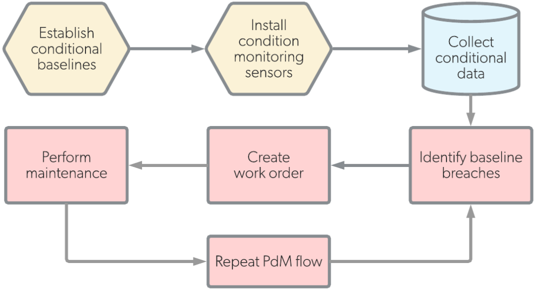
Manager 1

Exactly - either we're over-maintaining or under-maintaining, and neither side of the scale feels right. What we really need is a way to know when something's actually about to go wrong, not just guess.

Maintenance – Two alternative workflows



Preventive maintenance


Establish conditional baselines
Install condition monitoring sensors
Collect conditional data
Identify baseline breaches
Create work order
Perform maintenance
Repeat PdM flow

Predictive maintenance

Reference: `https://ukeep.com`

Summary: Insights from the Managers' Discussion I

Key Observations

- CNC machines are critical assets, operating under continuous and high-precision conditions.
- Unplanned downtime has direct consequences on production efficiency, costs, and delivery commitments.
- Traditional maintenance strategies (reactive or scheduled) are no longer sufficient to prevent unexpected failures.
- Environmental and operational factors accelerate wear and reduce component lifespan.
- Although CNC machines generate rich sensor data, current practices do not fully exploit it for predictive insights.

Summary: Insights from the Managers’ Discussion II

Strategic Implication

- The discussion highlights the need for a transition toward **predictive maintenance**, using data-driven methods to anticipate failures.
- Investing in advanced analytics and monitoring systems will increase reliability, optimize maintenance resources, and protect production continuity.

Why We Need Predictive Maintenance

- CNC machines are the backbone of our production lines
  - operate continuously under high precision requirements.
- Despite preventive schedules, we still face:
  - Unexpected failures causing production halts
  - Missed delivery deadlines
  - Rising maintenance and repair costs
- Current approach: mostly **reactive** or **time-based maintenance**.
- Opportunity: transition to a **predictive, data-driven strategy**.

Expected Benefits of Predictive Maintenance

1. Minimize Unplanned Downtime

Predict failures before they occur, allowing maintenance teams to act proactively and schedule interventions during planned stoppages.

2. Optimize Maintenance Operations

Move from reactive or scheduled maintenance to **condition-based actions**. Prioritize tasks according to actual risk and machine condition.

3. Reduce Costs

Prevent major breakdowns by detecting small anomalies early. Improve team efficiency and extend component lifespan.

Impact on Productivity and Reliability

- Maintain CNC machines at maximum operational time.
- Reduce idle periods and rework caused by unexpected breakdowns.
- Enhance production reliability for tight delivery schedules.
- Improve customer satisfaction through consistent delivery performance.

In short:

Predictive maintenance transforms maintenance from a cost center into a **strategic enabler of efficiency and reliability**.

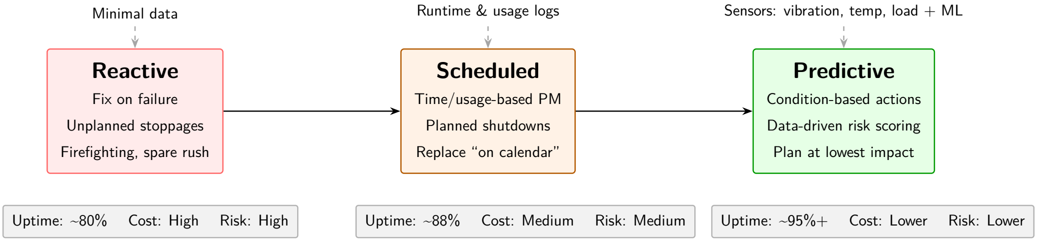
Next Steps and Recommendation

- **Step 1:** Launch a pilot project on a selected CNC production line.
- **Step 2:** Integrate real-time sensor data (vibration, temperature, load) for predictive analytics.
- **Step 3:** Evaluate benefits in uptime, cost, and reliability after 3 – 6 months.
- **Step 4:** Plan plant-wide deployment based on pilot results.

Our Recommendation

Investing in predictive maintenance is a strategic move toward **smart manufacturing**, aligning with our goals of cost reduction, reliability, and continuous improvement.

Maintenance Maturity



Minimal data

Reactive

Fix on failure

Unplanned stoppages

Firefighting, spare rush

Uptime: ~80% Cost: High Risk: High

Runtime & usage logs

Scheduled

Time/usage-based PM

Planned shutdowns

Replace “on calendar”

Uptime: ~88% Cost: Medium Risk: Medium

Sensors: vibration, temp, load + ML

Predictive

Condition-based actions

Data-driven risk scoring

Plan at lowest impact

Uptime: ~95%+ Cost: Lower Risk: Lower

Bottom line

Transitioning to **predictive** maintenance increases availability, reduces emergency repairs, and aligns interventions with production windows

Outline

1. Business Understanding 4
   - Informal discussion 5
   - Insights from the informal discussion 12
2. Data understanding 19
   - Example data 23
3. Data Preparation 32
4. Modelling 41
5. Evaluation 49
6. Deployment 58
7. Final Notes 67
8. 71

Data Collection: Overview

- Data collection is critical for building an effective predictive maintenance system
- The case study relies on various data sources to monitor machine performance and predict failures

Data Sources I

- Sensor Data
  - Collected from embedded sensors on CNC machines
  - Examples of metrics:
    - Vibration levels
    - Temperature readings
    - Pressure levels
    - Motor current usage
  - Recorded at high frequency (e.g., every second) during machine operation
- Maintenance Logs
  - Historical data detailing:
    - Repair activities
    - Component replacements
    - Failure events

Data Sources II

- Operational Data
  - Includes:
    - Machine workload levels
    - Runtime hours
    - Environmental conditions (e.g., humidity, ambient temperature)
- Quality Control Reports
  - Tracks defect rates in manufactured parts
  - Serves as an indirect indicator of machine performance issues

Example of Machine Details

- Machine ID: CNC-MILL-07
- Location: Plant 3 – Workshop A
- Date/Time of Failure: 2025-10-05, 14:37
- Operator: Marco Rossi

Failure Description

- Type of Failure: Unexpected spindle stoppage
- Symptoms:
  - Spindle RPM dropped from 12000 to 0 within 4 sec
  - Spindle temperature exceeded 85 C
  - Vibration spike (2.1 mm/s to 6.7 mm/s)
  - Audible screeching noise

Machine Parameters at Failure

Parameter	Value	Normal Range	Notes
Spindle Speed	12000 -> 0 RPM	5000–15000	Sudden drop
Spindle Temp	85 C	< 60 C	Overheating
Vibration RMS	6.7 mm/s	< 2.5 mm/s	Excessive
Motor Current	48 A	25–35 A	Overload
Tool Load	125 %	70–90 %	Excessive force

Failure Dataset Example

Record ID	Timestamp	Spindle RPM	Temp (C)	Vib RMS (mm/s)	Current (A)	Tool Load (%)	Failure
1	2025-10-05 14:37	12000	85	6.7	48	125	Yes
2	2025-10-05 14:15	11000	58	2.0	30	85	No
3	2025-10-05 13:50	9500	62	3.4	37	95	Warning
4	2025-10-05 13:20	8000	55	2.2	28	80	No
5	2025-10-05 12:55	10000	70	4.1	40	105	Yes

Root Cause Analysis

- Probable Cause: Spindle bearing degradation
- Contributing Factors:
    - Operation without scheduled lubrication
    - Tool misalignment in recent setup
    - High-load milling of hardened steel

Maintenance Actions

- Machine shut down immediately
- Cooling system checked (functional)
- Bearings inspected – wear and debris found
- Bearings replaced, lubrication restored
- Machine tested at reduced load, normal operation resumed

Recommendations

- Real-time vibration monitoring with accelerometer and FFT
- Set warning thresholds:
  - Vibration RMS > 3.0 mm/s
  - Spindle Temp > 65 C
- Automated lubrication every 200 hours
- Predictive model using historical sensor data

Challenges in Data Collection

- Diverse structure
  - e.g. quality control reports vs sensors data
- Noisy Sensor Data
  - High-frequency data often contains noise due to environmental interference or faulty sensors
- Missing Values
  - Occasional connectivity issues result in gaps in data streams
- Imbalanced Dataset
  - Failures are rare compared to normal operation
  - Imbalance makes it harder for predictive models to detect failure patterns

Significance of Collected Data

- Holistic View
  - Combining sensor, maintenance, operational, and quality data provides a complete picture of machine health
- Key Insights
  - Sensor data identifies real-time anomalies
  - Historical logs provide trends and failure patterns
  - Quality control links machine performance to product defects
- Data-Driven Decisions
  - Enables accurate failure predictions and optimized maintenance scheduling

Outline

1 Business Understanding 4
- Informal discussion 5
- Insights from the informal discussion 12

2 Data understanding 19
- Example data 23

3 Data Preparation 32

4 Modelling 41

5 Evaluation 49

6 Deployment 58

7 Final Notes 67

8 71

Data Preparation: Overview

- Preparing data for predictive maintenance in CNC machines involves cleaning, organizing, and enhancing data to extract actionable insights
- The process ensures raw sensor, maintenance, and operational data is suitable for machine learning and predictive analytics

Data Cleaning for CNC Machines

- Outlier Removal
  - Sensor data often contains abnormal readings caused by noise or transient events
  - Statistical methods like z-score analysis are used to identify and remove outliers
- Handling Missing Data
  - Causes of missing data:
    - Sensor malfunctions
    - Connectivity issues
  - Methods for imputation:
    - Mean/Median Imputation Suitable for stable, continuous variables
    - k-Nearest Neighbors (k-NN) Leverages similarity among instances for accurate estimation

Feature Selection for CNC Machines

- Key Sensor Features
  - Vibration statistics (e.g., mean, variance, skewness)
  - Temperature patterns (e.g., peak and trend analysis)
- Operational Features
  - Machine workload
  - Runtime and idle time metrics
- Aggregated Features
  - Cross-sensor interactions (e.g., vibration changes correlated with temperature spikes)

Feature Engineering for CNC Machines

- Sensor-Based Features
  - Extract key statistics:
    - Mean, standard deviation, and range of vibration signals
    - Temperature gradients over time
- Domain-Specific Features
  - Derived using knowledge of CNC machine operations:
    - Rate of spindle speed variation
    - Frequency of abnormal motor current spikes
- Cross-Feature Aggregation
  - Combine data from multiple sensors to uncover complex patterns
  - Example:
    - Correlating high temperature with rapid vibration changes to predict bearing wear

Normalization for CNC Machines

- Purpose
  - Ensure data from different sensors (e.g., temperature in °C, vibration in m/s^2) is on a comparable scale
- Methods
  - Min-Max Normalization
    - Scales data to a [0, 1] range
  - Z-Score Normalization
    - Standardizes data to have a mean of 0 and a standard deviation of 1

Dimensionality Reduction for CNC Data

- Need
  - High-frequency sensor data often results in high dimensionality
  - Reducing dimensions improves computational efficiency and removes redundant features
- Technique: Principal Component Analysis (PCA)
  - Captures the most critical information by transforming data into principal components
  - Retains significant variance while discarding noise

Final Prepared Dataset

- Integrated and Cleaned Data
  - Combines historical maintenance logs, operational data, and sensor data
- Key Features
  - Statistical metrics (e.g., mean, standard deviation)
  - Domain-specific insights (e.g., heat dissipation rate)
  - Aggregated cross-sensor indicators (e.g., combined vibration and temperature trends)
- Normalized and Reduced
  - Scaled and transformed data ready for predictive modeling

Significance of Data Preparation

- Improved Prediction Accuracy
  - Clean and enriched data leads to better model performance
- Operational Efficiency
  - Focused on relevant features, reducing unnecessary computational overhead
- Actionable Insights
  - Enables early detection of potential failures in CNC machines

Outline

1 Business Understanding 4
- Informal discussion 5
- Insights from the informal discussion 12

2 Data understanding 19
- Example data 23

3 Data Preparation 32

4 Modelling 41

5 Evaluation 49

6 Deployment 58

7 Final Notes 67

8 71

Modeling for Predictive Maintenance: Overview

- Predictive maintenance models aim to forecast failures or predict the remaining useful life (RUL) of CNC machines
- The approach combines machine learning techniques with domain-specific knowledge of CNC operations

Types of Predictive Models I

- Classification Models
  - Objective: Predict whether a failure will occur within a specified time frame
  - Example algorithms:
    - Logistic Regression
    - Support Vector Machines (SVM)
    - Random Forest
- Regression Models
  - Objective: Estimate the RUL of machine components
  - Example algorithms:
    - Linear Regression
    - Gradient Boosted Trees (e.g., XGBoost, LightGBM)

Types of Predictive Models II

- Anomaly Detection Models
    - Objective: Identify abnormal operating conditions indicating potential failure
    - Example algorithms:
        - Autoencoders
        - Isolation Forest
        - DBSCAN Clustering

Model Training Process

- Data Splitting
  - Dataset divided into training, validation, and test sets
  - Ensures robust performance evaluation
- Handling Class Imbalance
  - Failures are rare compared to normal operations
  - Techniques used:
    - Oversampling minority class using SMOTE (Synthetic Minority Oversampling Technique)
    - Undersampling majority class
- Cross-Validation
  - k-Fold Cross-Validation ensures model generalization

Evaluation Metrics

- Classification Metrics
  - Accuracy, Precision, Recall, and F1-Score for binary failure prediction
  - ROC-AUC for assessing overall performance
- Regression Metrics
  - Mean Absolute Error (MAE)
  - Root Mean Square Error (RMSE)
  - R^2 Score
- Anomaly Detection Metrics
  - Precision-Recall Curve for imbalanced datasets
  - Mean Squared Reconstruction Error for autoencoders

Advanced Techniques

- Deep Learning Models
  - Recurrent Neural Networks (RNNs)
    - Capture temporal patterns in sequential sensor data
  - Convolutional Neural Networks (CNNs)
    - Analyze sensor data as images (e.g., spectrograms of vibration signals)
- Hybrid Approaches
  - Combine traditional machine learning with deep learning for feature extraction and prediction
- Transfer Learning
  - Leverage pretrained models for specific failure scenarios

Deployment of Predictive Models

- Real-Time Integration
  - Models deployed on edge devices for real-time failure prediction
  - Data pipelines established for continuous sensor data monitoring
- Periodic Retraining
  - Models updated with new operational and failure data
  - Ensures adaptability to evolving machine conditions
- Integration with Maintenance Systems
  - Predictive outputs trigger automated maintenance scheduling
  - Reduces human intervention and response time

Outline

1 Business Understanding 4
- Informal discussion 5
- Insights from the informal discussion 12

2 Data understanding 19
- Example data 23

3 Data Preparation 32

4 Modelling 41

5 Evaluation 49

6 Deployment 58

7 Final Notes 67

8 71

Evaluation for Predictive Maintenance: Overview

- Evaluation ensures the predictive model's effectiveness and reliability in identifying failures or estimating Remaining Useful Life (RUL)
- It involves measuring performance against specific metrics tailored to the model's objectives

Evaluation Metrics: Classification Models

- Accuracy
  - Suitable for balanced datasets but less informative for imbalanced cases
- Precision
  - Focuses on the fraction of predicted failures that are correct
  - Important when false positives are costly (e.g., unnecessary maintenance)
- Recall
  - Measures the proportion of actual failures that are correctly predicted
  - Critical when missing a failure is unacceptable
- F1-Score
  - Balances false positives and false negatives
- ROC-AUC
  - Evaluates the trade-off between true positive and false positive rates
  - Suitable for comparing different classification models

Evaluation Metrics: Regression Models

- Mean Absolute Error (MAE)
  - Measures the average absolute difference between predicted and actual RUL
  - Easy to interpret and sensitive to large errors
- Root Mean Square Error (RMSE)
  - Penalizes large errors more heavily than MAE
  - Suitable when large deviations are particularly undesirable
- R^2 Score
  - Indicates the proportion of variance in RUL explained by the model
  - Higher values signify better model performance

Evaluation Metrics: Anomaly Detection Models

- Precision-Recall Curve
  - Evaluates performance in detecting rare failure events
  - Focuses on balancing false positives and true positives in imbalanced datasets
- Reconstruction Error
  - Used for models like autoencoders
  - Measures how well the model reconstructs normal behavior, flagging deviations as anomalies

Cross-Validation for CNC Machines

- Purpose
  - Ensures models generalize well to unseen data
- k-Fold Cross-Validation
  - Dataset is split into k subsets (folds)
  - Each fold is used as a test set while the others are used for training
  - Helps assess model stability and reliability
- Time-Based Validation
  - For sequential sensor data, ensures training data precedes test data
  - Prevents data leakage and ensures realistic evaluation

Interpretation of Results

- Threshold Tuning
  - Adjust decision thresholds based on evaluation metrics
  - Trade-offs:
    - Higher recall often reduces precision
    - Balance depends on operational priorities
- Root Cause Analysis
  - Evaluate feature importance to identify failure drivers
  - Helps optimize CNC machine operations
- Model Comparisons
  - Compare multiple models using consistent metrics and validation methods
  - Select the model with the best trade-off between accuracy, complexity, and interpretability

Challenges in Evaluation

- Imbalanced Datasets
    - Failure events are rare, leading to biased accuracy
    - Metrics like precision, recall, and F1-score are preferred
- Dynamic Conditions
    - Machine operating conditions vary over time
    - Continuous retraining and re-evaluation are required
- Complex Failure Patterns
    - Subtle anomalies may be missed by simple models
    - Advanced evaluation metrics (e.g., precision-recall curves) provide deeper insights

Significance of Evaluation

- Ensures Reliability
  - Models are tested for robustness under real-world scenarios
- Informs Deployment Decisions
  - Helps decide whether a model is ready for real-time integration
- Supports Continuous Improvement
  - Identifies weaknesses to guide model tuning and retraining

Outline

1 Business Understanding 4
- Informal discussion 5
- Insights from the informal discussion 12

2 Data understanding 19
- Example data 23

3 Data Preparation 32

4 Modelling 41

5 Evaluation 49

6 Deployment 58

7 Final Notes 67

8 71

Deployment for Predictive Maintenance: Overview

- Deployment involves integrating predictive maintenance models into CNC machine workflows
- It ensures real-time failure prediction and supports proactive maintenance decisions
- Key steps include infrastructure setup, integration with existing systems, and model monitoring

Infrastructure Requirements

- Edge Computing
  - Deploy models on local devices near CNC machines
  - Reduces latency for real-time predictions
- Cloud Integration
  - Centralized storage and processing for large-scale data analytics
  - Supports periodic retraining and model updates
- Data Pipelines
  - Establish automated pipelines for continuous data collection, preprocessing, and prediction
  - Ensure data security and compliance with industry standards

Real-Time Model Deployment

- Predictive Models at the Edge
  - Models predict machine health based on live sensor data
  - Outputs are delivered to operators or maintenance systems in real time
- Integration with Machine Control Systems
  - Alerts generated by models trigger actions:
    - Maintenance scheduling
    - Emergency shutdown to prevent damage
- Latency Optimization
  - Ensure prediction speed meets real-time requirements
  - Use optimized algorithms and hardware accelerators

Model Monitoring and Updates

- Performance Monitoring
  - Continuously evaluate prediction accuracy in real-world conditions
  - Metrics to monitor:
    - False positives triggering unnecessary maintenance
    - Missed failures causing downtime
- Drift Detection
  - Identify changes in data distribution due to new operating conditions or equipment upgrades
  - Retrain models periodically to maintain accuracy
- Feedback Loops
  - Incorporate operator feedback and maintenance outcomes to refine models

Integration with Maintenance Systems

- Automated Maintenance Triggers
  - Predictive models send alerts to maintenance management systems
  - Systems schedule maintenance based on severity and priority
- Downtime Minimization
  - Predictions align maintenance with planned downtimes
  - Reduces unexpected halts in production
- User-Friendly Dashboards
  - Visualize real-time machine health and predictions
  - Provide actionable insights to operators and engineers

Scalability and Adaptability

- Scalable Solutions
  - Models are designed to handle increasing numbers of CNC machines and sensors
  - Cloud platforms support horizontal scaling
- Adaptability to New Machines
  - Transfer learning enables rapid adaptation to new CNC models
  - Fine-tune existing models with minimal retraining
- Customizable Pipelines
  - Allow for easy addition or modification of sensors and features
  - Accommodates changes in machine configurations

Challenges in Deployment

- Data Security
  - Ensure compliance with industrial data protection regulations
  - Implement encryption and secure access controls
- Model Reliability
  - Validate models under varied operating conditions
  - Handle edge cases effectively
- Cost of Infrastructure
  - Balance the trade-off between edge and cloud resources
  - Optimize investments in hardware and computational resources

Expected Benefits after Deployment

- Reduced Downtime
  - Predictive alerts prevent unexpected machine failures
- Cost Savings
  - Minimizes unnecessary maintenance and part replacements
- Enhanced Efficiency
  - Enables operators to focus on critical tasks, improving overall productivity

Outline

1. Business Understanding 4
   - Informal discussion 5
   - Insights from the informal discussion 12
2. Data understanding 19
   - Example data 23
3. Data Preparation 32
4. Modelling 41
5. Evaluation 49
6. Deployment 58
7. **Final Notes 67**
8. 71

Impact (rough estimate)

- Cost Savings
  - Reduced downtime by 25%
  - Maintenance costs decreased by 15%
- Improved Productivity
  - Production reliability increased by 20%
- Employee Efficiency
  - Focus shifted to high-priority tasks instead of routine inspections

Conclusion

- Data mining techniques proved effective for predictive maintenance in CNC machines
- Combined feature engineering, machine learning models, and real-time monitoring reduced costs and improved efficiency
- Approach is scalable to other industrial environments (e.g., oil refineries, logistics)

Future Work

- Integration with Digital Twins
  - Simulate machine behavior for more robust predictions
- Enhanced Models
  - Incorporate Reinforcement Learning for adaptive maintenance
- Scalability
  - Deploy solutions across diverse facilities

"The only way to discover the limits of the possible is to go beyond them into the impossible."

Arthur C. Clarke

Bibliography

- Shearer, C. (2000).
  The CRISP-DM model: The new blueprint for data mining.
  Journal of Data Warehousing, 5:13–22.

- Wirth, R. and Hipp, J. (2000).
  CRISP-DM: Towards a standard process model for data mining.
  Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining.