Product Quality Output Measurement for Preventive Maintenance on Computer Numerical Control (CNC) Machines at an Electronic Manufacturing Industry

Aizat Haikal Bin Apandi[1*], Adam Shariff[1], Kamarulzaman Mahmad Khairai[2]

Centre Science and Mathematics, University Malaysia Pahang As-Sultan Abdullah, Pahang, Malaysia
Kuliyyah of Engineering, International Islamic University Malaysia, Selangor, Malaysia
* Correspondence: azthkl05@gmail.com

Received: 8 October 2024
Revised: 19 December 2024
Accepted: 30 December 2024

Citation:
Apandi, A. H. B., Shariff, A., & Khairai, K. M. (2024). Product Quality output measurement for preventive maintenance on computer numerical control (CNC) machines at an electronic manufacturing industry. *QOMARUNA Journal of Multidisciplinary Studies, 2*(1), 1–19.

ABSTRACT
Computer Numerical Control (CNC) machines remove material from a blank or workpiece using digital controls to produce custom-designed parts. Maintaining their accuracy and precision under challenging conditions after long-term usage is crucial. This study aims to evaluate CNC product quality using Overall Equipment Effectiveness (OEE) and enhance long-term performance through data-driven approaches. The method of this study focuses on analyzing scrap rate data, employing a u-chart to monitor stability, and applying machine learning regression models—K-Nearest Neighbour (KNN) and Random Forest (RF)—to forecast scrap rates. These forecasts help identify when preventive maintenance is necessary, preserving machine precision over time. This study also applied visualization of results with Microsoft Power BI to enhance data interpretation, aiding quick responses to potential problems. Results indicate that RF outperforms KNN in predicting scrap rates. Stacking these models further improves accuracy, offering a more reliable decision-making tool for anticipating quality issues. By detecting anomalies early, manufacturers can implement timely maintenance, minimizing downtime and prolonging CNC machine lifespan. In conclusion, integrating scrap rate analysis, statistical process control, and advanced machine learning techniques can maintain product quality and reduce inaccuracies. Companies should include more proactive maintenance planning by employing better forecasting.

Keywords: *Computer Numerical Control (CNC), Overall Equipment Effectiveness (OEE), preventive maintenance, quality improvement, machine learning,*

ABSTRAK
Mesin *Computer Numerical Control* (CNC) menghilangkan material dari sebuah benda kerja dengan menggunakan kontrol digital untuk menghasilkan komponen yang dirancang secara khusus. Penting untuk menjaga akurasi dan presisi mesin ini di bawah kondisi yang menantang setelah penggunaan jangka panjang. Penelitian ini bertujuan untuk mengevaluasi kualitas produk CNC menggunakan Overall Equipment Effectiveness (OEE) dan meningkatkan kinerja jangka panjang melalui pendekatan berbasis data. Metode penelitian ini berfokus pada analisis laju scrap, penggunaan *u*-chart untuk memantau stabilitas, serta penerapan model regresi *machine learning*—K-Nearest Neighbour (KNN) dan Random Forest (RF)—untuk memprediksi laju scrap. Prediksi tersebut membantu mengidentifikasi waktu yang tepat untuk melakukan pemeliharaan pencegahan, sehingga presisi mesin dapat tetap terjaga seiring waktu. Penelitian ini juga memanfaatkan visualisasi hasil menggunakan Microsoft Power BI untuk meningkatkan interpretasi data dan memfasilitasi respons cepat terhadap potensi masalah. Hasil penelitian menunjukkan bahwa RF memiliki kinerja lebih baik dibandingkan KNN

dalam memprediksi laju scrap. Penggunaan stacking pada model-model tersebut lebih lanjut meningkatkan akurasi, sehingga memberikan alat pengambilan keputusan yang lebih andal dalam mengantisipasi masalah kualitas. Dengan mendeteksi anomali secara dini, produsen dapat melakukan pemeliharaan tepat waktu, meminimalkan waktu henti, serta memperpanjang umur operasional mesin CNC. Kesimpulannya, integrasi analisis laju scrap, pengendalian proses statistik, dan teknik pembelajaran mesin yang canggih dapat secara efektif menjaga kualitas produk serta mengurangi ketidakakuratan. Perusahaan sebaiknya mengadopsi perencanaan pemeliharaan yang lebih proaktif dengan memanfaatkan peramalan yang lebih baik.

Keywords: *Computer Numerical Control (CNC), Overall Equipment Effectiveness (OEE), preventive maintenance, perbaikan kualitas, machine learning*

1. Introduction

Nowadays, manufacturing companies supplying electronic components recognize the critical importance of product quality to maintain customer satisfaction and competitiveness. Ensuring product quality begins with maintaining precise, reliable production equipment. Preventive maintenance plays a key role in improving machines' Overall Equipment Effectiveness (OEE), thereby enhancing performance, availability, and product quality. Moghaddam (2015) defines preventive maintenance as scheduled activities over a planning horizon to extend a system's useful life and maintain responsiveness, ultimately making it more reliable and available.

Within manufacturing, computer numerical control (CNC) machines form the backbone of precision-driven production processes. CNC technology removes material layers from a workpiece using digitally controlled machine tools, enabling high-accuracy, custom-designed parts. Over extended operation, maintaining precision under challenging conditions becomes essential to ensure product quality (Liu, et al, 2022). Preventive maintenance strategies for CNC machines help preserve this accuracy and reduce product scrap. This benefits both the manufacturer's profitability and reputation.

The transformative potential of predictive maintenance has been extensively demonstrated in prior research. Shahin et al. (2023) investigated over 20 fault detection models using Machine Learning (ML), Deep Learning (DL), and Deep Hybrid Learning (DHL) to minimize manufacturing downtime, achieving significant accuracy improvements in early failure detection. Traini (2019) introduced a machine learning framework for predictive maintenance in milling, leveraging Industrial IoT and AI technologies to enhance real-time monitoring and defect detection. Similarly, Paolanti (2018) employed Random Forest models within a Machine Learning architecture to improve system reliability and prevent unexpected equipment failures. These studies emphasize the effectiveness of advanced analytics in optimizing maintenance practices, reducing defects, and improving reliability.

Despite these advancements, case-specific investigations remain critical, particularly for established companies like "Company T," an electronic manufacturer located in Pahang, Malaysia, serving the aerospace, automotive, medical, and industrial sectors. Although the company has operated for over two decades, it has yet to implement predictive maintenance practices. Addressing this gap offers an opportunity to tailor state-of-the-art methodologies to real-world challenges, ensuring consistent product quality and operational efficiency.

This study aims to investigate the application of preventive maintenance and predictive tools in improving CNC machine quality control, particularly for air coils, which are critical components for automotive and industrial clients. By analyzing measurement data, employing advanced statistical and machine learning approaches, and implementing systematic maintenance strategies, this research aims to minimize product scrap, enhance OEE, and improve customer satisfaction while providing practical solutions tailored to the specific operational conditions of "Company T."

.

2. Literature Review

2.1 Computer Numerical Control (CNC) Machine

Computer Numerical Control (CNC) machines reflect a region's industrial development and are central to advancing aerospace, military, automotive, and electronic manufacturing. They have long played a key role in national competitiveness. Among their performance measures, machine accuracy is paramount, as a part's precision ultimately depends on the machine's machining accuracy (Liu et al., 2022). Their accuracy, which relates to how closely finished parts match design specifications, is a key performance measure. As noted in "Monitoring of CNC Machine Tool Accuracy," proper maintenance and adherence to certified standards (ISO, DIN, EN) ensure machines meet precision requirements.

Perfect accuracy is rarely attainable. Holub's "Geometric Error Compensation of CNC Machine Tool" highlights that actual part dimensions seldom match nominal design values due to process variations. Finishing operations, however, can bring measurements closer to the target. Manufacturers enhance product precision and maintain high-quality standards by emphasizing quality measures and reducing output errors.

Figure 1. Computer Numerical Control (CNC) machine (Documentation of Company "T")

2.2. Overall Equipment Effectiveness (OEE)

Overall Equipment Effectiveness (OEE) is an excellent indicator for measuring sustainability improvements relative to a company's initial operational state. As a key performance indicator (KPI), OEE reflects not only a machine or system's performance, but also the effectiveness of the personnel responsible for its maintenance (Haddad et al., 2021). It integrates aspects of operation, maintenance, and resource management (Tsarouhas, 2020).

Although OEE could not reveal the exact causes of inefficiencies, it helps categorize areas needing improvement (Tsarouhas, 2020). Enhancing OEE can boost production capacity, improve product quality, reduce downtime, and increase overall system efficiency. According to P. H. Tsarouhas (2020), OEE is determined by three critical parameters: availability, performance, and quality. Figure 2 illustrates how OEE is calculated based on these factors.

Measuring OEE effectively evaluates the efficiency of a single machine or an integrated manufacturing system (Dewi, et al., 2020). Widely recognized as a measure of internal efficiency, OEE reflects a machine's true value-added output. It helps identify equipment-related losses to improve overall asset performance and reliability.

These losses can be grouped into six major categories (Tsarouhas, 2020): equipment failure, setup and adjustment delays, minor stoppages and idle times, speed reductions, defects or rework, and reduced performance (Tsarouhas, 2019). By addressing these losses, improving OEE results in fewer breakdowns, reduced idle times, lower defect rates, and fewer workplace accidents. It also boosts productivity, optimizes processes, encourages workforce involvement, increases profits through cost savings, enhances customer satisfaction, and raises employee morale and confidence. Figure 3 illustrates these six major categories.

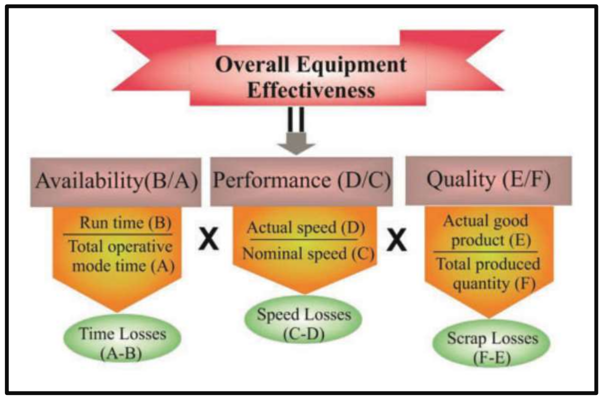


Figure 2. OEE Calculation (Tsarouhas, 2020)

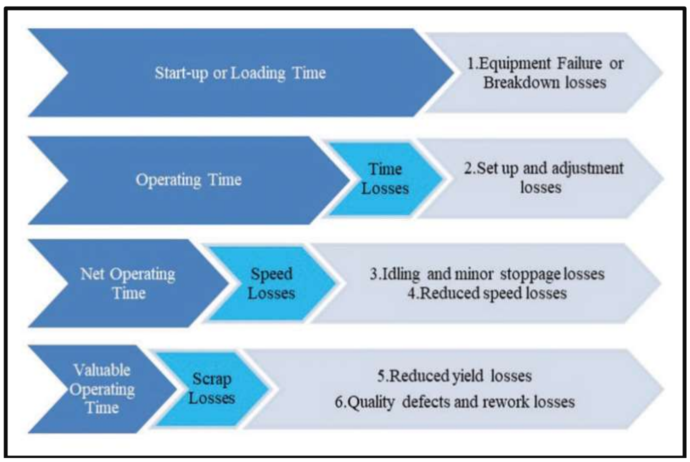


Figure 3. Six Major OEE (Tsarouhas, 2019)

Quality

Quality is one of the three factors used to determine OEE. The quality rate measures how many non-defective products a machine produces during its run time (Dewi, et al., 2020). It evaluates the percentage of satisfactory units produced, accounting for defects, scrap, and rework that affect final product quality. To calculate quality, count the number of units meeting required standards and divide by the total number produced while operating (Tsarouhas, 2020):

$$\text{Quality}=\frac{\text{Good Units}}{\text{Total Units}}\times 100\% \tag{1}$$

Improving quality involves implementing quality control checks, enhancing equipment reliability, and providing better training and procedures. These measures raise the OEE score by reducing scrap and rework, increasing customer satisfaction, and improving overall product quality.

Performance

Performance is one of the three OEE components, measuring how effectively equipment operates relative to its maximum potential output (Singh, et al., 2021). It accounts for speed losses, minor stops, and slow cycles that reduce the rate at which products are produced, even if the equipment is fully available. Performance is calculated by comparing the total products produced to what could have been produced under optimal conditions (Dewi, et al., 2020):

$$\text{Performance} = \frac{\text{Actual Output}}{\text{Maximum Possible Output}} \times 100\% \tag{2}$$

Increasing performance can involve maximizing production rates, minimizing downtime between cycles, streamlining changeover times, and providing better training and procedures. By enhancing performance, you can increase both the equipment's productivity and its OEE score.

Availability

Availability is the third component of OEE and measures the percentage of time equipment is ready for use (Haddad, et al, 2021). It accounts for both scheduled and unscheduled downtime, as well as any other factors that prevent the equipment from operating at full capacity. Scheduled downtime includes planned maintenance, changeovers, and other predefined events requiring the equipment to be offline. Unplanned downtime covers unexpected breakdowns, repairs, and incidents that halt normal operations. To determine availability, you must identify the total time the equipment ran productively and compare it to the planned production time during the measurement period. The formula for availability is:

$$\text{Availability} = \frac{\text{Operating Time - Downtime}}{\text{Operating Time}} \times 100\% \tag{3}$$

Operational time represents the total period the equipment is intended to run during production, while downtime is the total time it remains unavailable due to unexpected issues like breakdowns or maintenance. Improving availability involves preventive maintenance, reducing changeover time, enhancing equipment reliability, and implementing effective planning and scheduling. By increasing equipment availability, you boost overall productivity and the OEE score.

2.3. Preventive Maintenance

Preventive maintenance refers to scheduled activities performed at regular intervals to extend a system's useful life and keep it both productive and responsive (Moghaddam, 2015). These activities—such as inspections, cleaning, lubrication, adjustments, alignments, and component replacements—reduce a system's "effective age" and lower failure rates. Manufacturers typically provide recommended maintenance schedules to minimize unexpected failures over a machine's operational lifespan. While preventive maintenance improves reliability and availability, its designers must balance the costs of upkeep and replacement against the risks and expenses of unanticipated breakdowns.

Applied correctly, preventive maintenance ensures CNC machines remain efficient, producing parts that meet precise design specifications and maintaining high quality standards. By proactively addressing wear and tear, companies can reduce product defects, lower scrap rates, and improve their overall profitability. Systematic preventive maintenance planning also helps coordinate service intervals, allowing organizations to address issues before they arise and maintain optimal equipment performance. This strategic approach ultimately leads to more consistent product quality, increased customer satisfaction, and better long-term returns.

2.4. Predictive Maintenance

Predictive maintenance represents a proactive strategy that relies on data analysis and predictive modelling to anticipate equipment failures, enabling maintenance tasks to align with actual conditions rather than fixed schedules. Accordingly, organizations can reduce downtime and costs, a key driver for its growing adoption across multiple industries, including CNC machine facilities.

In a prior study, Lee et al. (2022) developed a predictive maintenance system for CNC machines that used Machine Learning (ML) models to forecast cutting tool loads. By anticipating periods of high stress, maintenance activities could be scheduled more effectively, minimizing downtime and extending tool life. However, the system's success depended heavily on accurate, high-quality sensor data and machine logs.

Expanding on this research, Soori et al. (2023) reviewed ML and AI techniques—such as regression models, support vector machines (SVMs), and artificial neural networks (ANNs)—applied to CNC predictive maintenance. They identified vast potential for ML and AI to revolutionize maintenance practices but noted challenges like machine performance variability and non-standardized data. The authors called for collaboration between data scientists and domain experts to address these issues and fully exploit ML and AI capabilities.

In another advancement, Ruiz Rodríguez et al. (2022) introduced multi-agent deep reinforcement learning (MADRL) for predictive maintenance in parallel machines. This adaptive system outperformed traditional methods by reducing downtime and extending machine lifespan. Nevertheless, its heavy computational demands may limit its applicability in certain environments.

3. Method

This study involves two main stages: first, constructing a control chart based on defects per unit produced by CNC machines; second, developing predictive models using various machine learning algorithms.

3.1. Control Chart

A u-chart, a key tool in statistical process control (SPC), monitors the number of defects per unit over time. It is particularly useful in quality control and manufacturing, where minimizing defects is essential for cost-effectiveness, customer satisfaction, and compliance with industry standards (Tang, et al., 2019).

To construct a *u*-chart, samples are collected at regular intervals, and the defect rate per unit is calculated for each sample. Care should be taken to ensure that the sample size (n) is representative, as variations in n affect the accuracy of the results. In cases where sample sizes differ, applying a weighted moving average can help maintain a consistent view of process performance. Equation (4) calculates the average defect rate per unit, which serves as the central line in the u-chart whereas Equation (5) provides the formulas for the Upper and Lower Control Limits (UCL and LCL). UCL and LCL reflect the acceptable range of variation for an in-control process. If data points fall outside these limits, it may indicate an out-of-control condition, prompting further investigation (Ieren, et al., 2020).

$$u = \frac{u}{n}$$
$$\bar{u} = \frac{\sum c}{\sum n} \tag{4}$$
$$\sigma p = \sqrt{\frac{\bar{u}}{n}}$$

where:
c: number of defects
k: number of lots
n: sample size

Upper Control Limit:
$$UCL_p = \bar{u} + 3\sigma p$$
Lower Control Limit:
$$LCL_p = \bar{u} + 3\sigma p \tag{5}$$

By applying control limits, one can determine whether a process is in control. A process is considered stable if plotted points remain within the control limits and follow a random pattern. Points that fall outside these limits or form non-random patterns suggest the presence of special causes that require investigation and correction. It is important to note that u-charts assume data follows a Poisson distribution and that defect rates are moderate. Extremely low defect rates may require large sample sizes, while very high rates may limit the chart's applicability.

Control charts like the u-chart are invaluable for monitoring process performance over time. They typically feature three lines: a central line representing the average and upper and lower control limits, often set at three standard deviations. Selecting an appropriate control chart is crucial, especially when dealing with attribute data such as defect counts.

After creating a control chart and interpreting its results, the next phase involves using the collected data to inform predictive models. In this investigation, two models—K-Nearest Neighbors (KNN) and Random Forest (RF)—will analyze data from the u-chart to predict parameters of interest (e.g., air coil measurements). These predictions help improve process understanding and guide informed decision-making.

3.2. Building Prediction Model

After creating the control chart, the next step is to use the collected information to train a predictive model for the parameter of interest—such as air coil measurements. In this study, two models will be employed: the K-Nearest Neighbors (KNN) algorithm and the Random Forest (RF) algorithm. Both will use the data obtained from the u-chart to generate predictions, ultimately enhancing our understanding of the underlying process.

K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a supervised learning algorithm applicable to both classification and regression tasks. In regression, KNN identifies the k nearest data points relative to a test data point and then predicts the output as the average value of these k neighbors. The prediction of KNN regression can be expressed as follows Rodriguez-Galiano, et al. (2015):

$$\hat{y} = \frac{1}{k} \sum_{x_i \in N(x)} y_i \tag{6}$$

Here, the algorithm assumes that all input variables are on the same scale and that the target variable is continuous. Choosing the optimal k is critical. A larger k tends to yield smoother predictions, while a smaller k may provide more flexibility. To determine the best k, it's often useful to experiment with different values and evaluate performance using cross-validation or other error metrics.

Random Forest

Random Forest regression is often employed for predicting continuous values rather than categorical outcomes. As an ensemble method, it combines multiple decision trees to produce a final prediction. Each decision tree is trained on a randomly selected subset of the training data (with replacement), and after all trees are trained, their predictions are averaged to yield the final result. Equation 7 shows the calculation of predicted value of y (Rodriguez-Galiano, et al., 2015):

The predicted output value for x is the average of the outputs predicted by all the trees:

$$RF(x) = \frac{1}{T} \sum_{t=1}^{T} \hat{y_t}(x) \tag{7}$$

Where:
x: input vector representing the set of features for which a prediction is to be made.
T: total number of individual decision trees that make up the Random Forest.

$\hat{y}_t(x)$: The prediction made by the t^{th} decision tree in the Random Forest for the input x

In this notation, $RF(x)$ is the predicted output value for the input vector x. Random Forest regression offers several advantages over other regression algorithms. It is less prone to overfitting, can efficiently handle high-dimensional data, trains and predicts relatively quickly, and can model non-linear relationships between input features and output values. Its performance can be fine-tuned by adjusting hyperparameters such as the number of trees, the maximum tree depth, and the number of features considered at each split. Optimal parameter values are often identified through cross-validation (Alquthami, et al., 2022). In this study, a RF regression model was applied to the air coil scrap data to generate predictions and meet the research objectives.

4. Results and Discussion

4.1 Results

Model Selection Analysis

Because the scope of the model is quite broad, a bar chart was used to pinpoint which models warrant closer attention. A threshold, derived from the overall scrap data, serves as a benchmark for selecting the models under review. This targeted approach simplifies the analysis, making it clearer and more aligned with the study's objectives.

Based on the production data, a threshold was established to identify which models require further analysis. As shown in Figures 6 and 7, seven machines exceeded this threshold—specifically machines 1, 2, 4, 5, 10, 11, and 16. The models involved are HA00-08464KLFTR, HA00-17359ALFTR, and HM73-10C300LFTR.

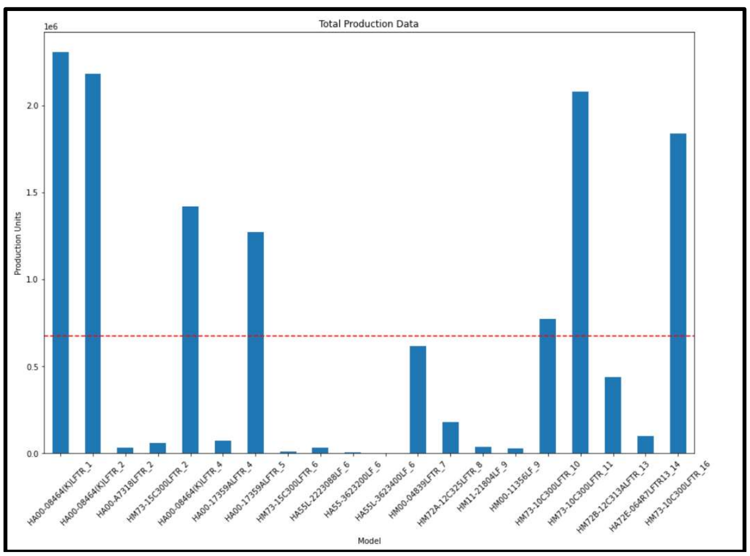
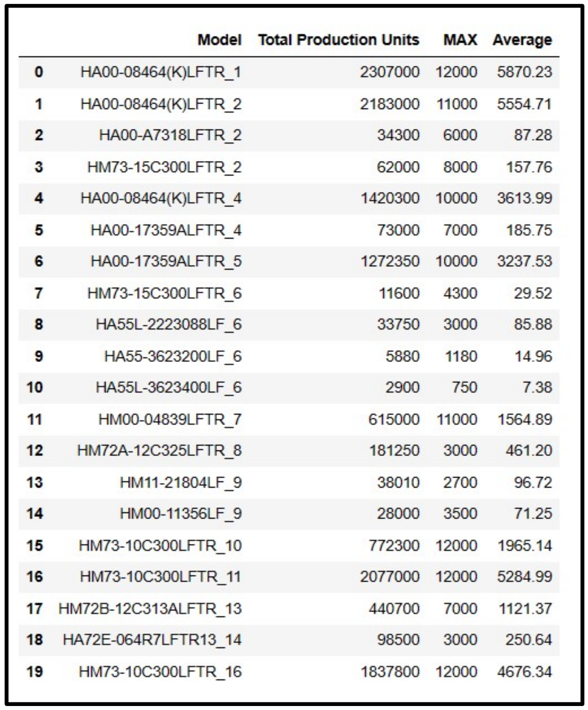


Figure 6. Threshold Productivity Model



	Model	Total Production Units	MAX	Average
0	HA00-08464(K)LFTR_1	2307000	12000	5870.23
1	HA00-08464(K)LFTR_2	2183000	11000	5554.71
2	HA00-A7318LFTR_2	34300	6000	87.28
3	HM73-15C300LFTR_2	62000	8000	157.76
4	HA00-08464(K)LFTR_4	1420300	10000	3613.99
5	HA00-17359ALFTR_4	73000	7000	185.75
6	HA00-17359ALFTR_5	1272350	10000	3237.53
7	HM73-15C300LFTR_6	11600	4300	29.52
8	HA55L-2223088LF_6	33750	3000	85.88
9	HA55-3623200LF_6	5880	1180	14.96
10	HA55L-3623400LF_6	2900	750	7.38
11	HM00-04839LFTR_7	615000	11000	1564.89
12	HM72A-12C325LFTR_8	181250	3000	461.20
13	HM11-21804LF_9	38010	2700	96.72
14	HM00-11356LF_9	28000	3500	71.25
15	HM73-10C300LFTR_10	772300	12000	1965.14
16	HM73-10C300LFTR_11	2077000	12000	5284.99
17	HM72B-12C313ALFTR_13	440700	7000	1121.37
18	HA72E-064R7LFTR13_14	98500	3000	250.64
19	HM73-10C300LFTR_16	1837800	12000	4676.34

Figure 7. Productivity by Model

U-Chart

The utilization of the U-chart proves to be a viable means of monitoring the caliber of output from a CNC apparatus via pre-emptive maintenance. The use of a chart aids in the detection of possible deficiencies before they develop into significant problems and contributes to the preservation of the machinery's output quality. The utilization of CNC machine assists in gauging the caliber of the generated output and guaranteeing its adherence to the prescribed criteria of acceptability. One limitation of this case study pertains to the temporal scope of the data, which is restricted to a one-year duration. It can be argued that the control chart is more efficacious than plotting data by date or daily. Through the graphical representation, it is possible to observe the progression of a certain procedure monthly. However, for greater specificity, a more detailed scrutiny can be carried out by analyzing the data daily. The monitoring process will follow as in section 3.8 to evaluate the u-chart. But the evaluation focuses on the data that occur anomalies and in control at that month.

According to the data in Figure 8, only in January and May did the process remain consistently within the established control limits. In contrast, significant volatility was observed during July and September, making process management challenging. To pinpoint the problematic machine within the HA00-08464LFTR model, it is necessary to determine whether machine 1, machine 2, or machine 4 is responsible for the observed instability.

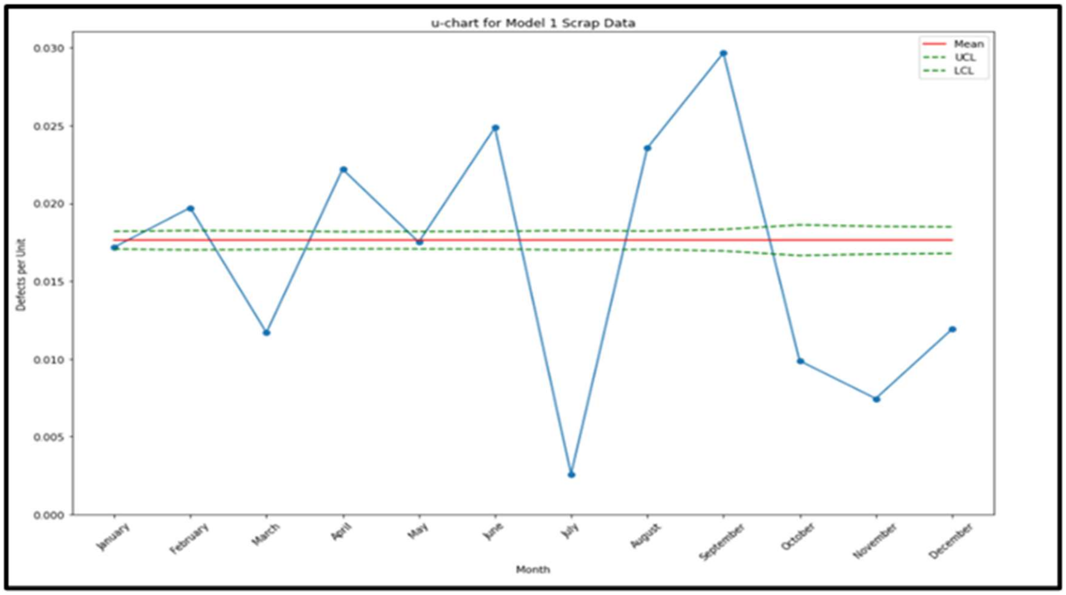


Figure 8. Control Chart for Model 1

According to the data in Figure 9, the process remained consistently within the established control limits during May, as indicated by the plot falling entirely within the designated control range. However, certain periods, notably July and September, showed greater variability, posing challenges for effective process management. Additionally, production ceased in November, coinciding with the conclusion of the Descriptive Analysis segment.

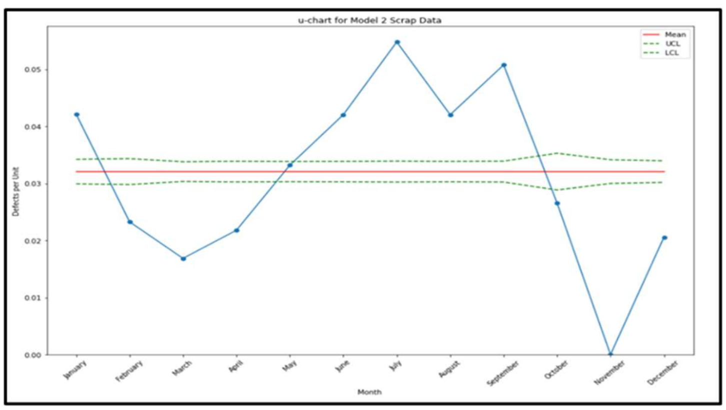


Figure 9. Control Chart Model 2

As depicted in Figure 10, the process control remained within established limits only during October, as both periods displayed plots confined to the specified range. In contrast, variability intensified in September, where a notable outlier emerged. Despite this anomaly, the process remained stable, though it fell outside the expected parameters. September's scrap rate reached the highest recorded peak, posing additional challenges to maintaining effective process oversight.

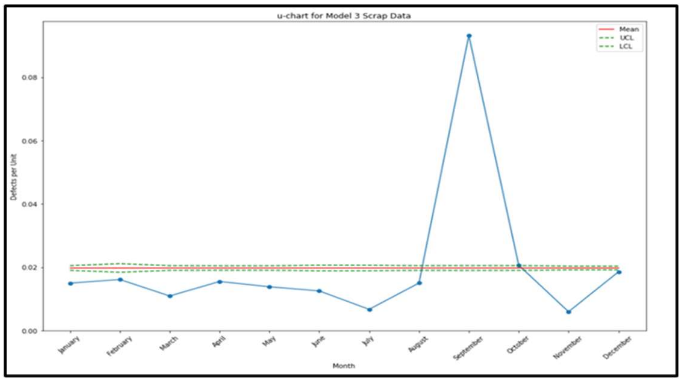


Figure 10. Control Chart Model 3

Regression Model

A forecasting analysis was conducted using data from 2022 on three models: Model 1 (HA00-08464LFTR), Model 2 (HA00-17359ALFTR), and Model 3 (HM73-10C300LFTR). The machine learning pipeline remains consistent for all three, differing only in column identifiers. Each model's data was split into training (80%) and testing (20%) sets to facilitate analysis.

Feature selection was then performed to identify the most relevant variables. By removing unnecessary, redundant, or less informative features, feature selection helps streamline the model, improve accuracy, reduce overfitting, and shorten training times.

Next, a hyperparameter tuning approach was applied to optimize the models (see Figure 11). Random search, a method for discovering the best set of hyperparameters, was employed. Hyperparameters are preset values that influence a model's training behavior and overall performance. By defining a parameter grid and conducting a grid search, the process finds the parameter combination that yields the best performance. Random search tuning was applied to both RF and KNN models, previously selected during the model selection phase.

```
# RandomizedSearchCV for hyperparameter tuning
param_dist_rf = {
    'n_estimators': [10, 50, 100, 200],
    'max_depth': [None, 10, 20, 30],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4]
}
random_search_rf = RandomizedSearchCV(rf, param_distributions=param_dist_rf, cv=5, n_iter=10)
```

```
# Define the hyperparameter grid for RandomizedSearchCV
param_dist_knn = {
    'n_neighbors': list(range(1, 6)),
    'weights': ['uniform', 'distance'],
    'algorithm': ['auto', 'ball_tree', 'kd_tree', 'brute'],
    'metric': ['euclidean', 'manhattan', 'minkowski']
}
random_search_knn = RandomizedSearchCV(knn, param_distributions=param_dist_knn, cv=5, n_iter=10)
```

Figure 11. Hyperparameter Tuning Code

Training Model

```
# Train the model
best_model_rf1 = random_search_rf1.best_estimator_
best_model_rf1.fit(X1_train_selected, y1_train)
```

```
RandomForestRegressor
RandomForestRegressor(max_depth=10, min_samples_split=10, n_estimators=10,
                      random_state=42)
```

```
# Train the model
best_model_knn1 = random_search_knn1.best_estimator_
best_model_knn1.fit(X1_train_selected, y1_train)
```

```
KNeighborsRegressor
KNeighborsRegressor(metric='euclidean', weights='distance')
```

Figure 12. Training Model Code

In this study, the chosen models are KNN and RF. Among them, RF is considered an ensemble model, while KNN is treated as a base model. The evaluation criteria include minimizing the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), while maximizing the correlation coefficient (*R*) (Chicco, et al., 2021). Based on these criteria, Table 1 indicates that the best-performing models are RF for Model 1 and Model 3, and KNN for Model 2.

Table 1. Evaluation for the best model

Keyword	Model	ML Model	RMSE	MAE	R
Model 1	HA00-08464LFTR	RF	0.0049	0.0024	0.898
		KNN	0.0100	0.0047	0.578
Model 2	HA00-17359ALFTR	RF	0.0457	0.0127	0.177
		KNN	0.0100	0.0156	0.523
Model 3	HM73-10C300LFTR	RF	0.008	0.0025	0.821
		KNN	0.013	0.0048	0.559

Although the model's performance was already established, further improvement can be achieved through stacking, an ensemble technique that layers multiple models to enhance predictive accuracy. By blending predictions from different models, stacking can bolster overall performance. As shown in Table 2, most models benefit from stacking except for Model 1, where the RF model alone still outperforms the stacked approach.

To validate these findings, actual and predicted values were compared. The actual data from January to March 2023 served as a reference, while the models were trained and tested exclusively on data from 2022 before forecasting and comparing the results against the 2023 actuals. This approach ensures a realistic assessment of model performance in a true production environment.

Figure 13. Stacking Method for Improvement of Model

Table 2. Evaluation of Model by Adding Stacking Method

Keyword	**Model**	**ML Model**	**RMSE**	**MAE**	**R**
Model 1	HA00-08464LFTR	RF	0.00489	0.00235	0.89849
		KNN	0.00997	0.00470	0.57845
		Stacking	0.00738	0.00315	0.76877
Model 2	HA00-17359ALFTR	RF	0.04568	0.01270	0.17666
		KNN	0.00997	0.01557	0.52287
		Stacking	0.03126	0.00886	0.61453
Model 3	HM73-10C300LFTR	RF	0.00812	0.00249	0.82134
		KNN	0.01276	0.00478	0.55867
		Stacking	0.00781	0.00333	0.83472

From Table 4, the stacking algorithm achieved the highest performance for Model 2 and Model 3, while Random Forest remained superior for Model 1. Figures 14, 15, and 16 compare actual and predicted values produced by RF and the stacking approach. Although RF reached higher peaks, the stacking predictions more closely matched the actual collected data. Unlike testing on historical data, these graphs demonstrate how the models perform on real, current data, making the assessment more practical for real-world forecasting scenarios. As a result, the stacking algorithm was selected to forecast the CNC scrap rate across the evaluated models.

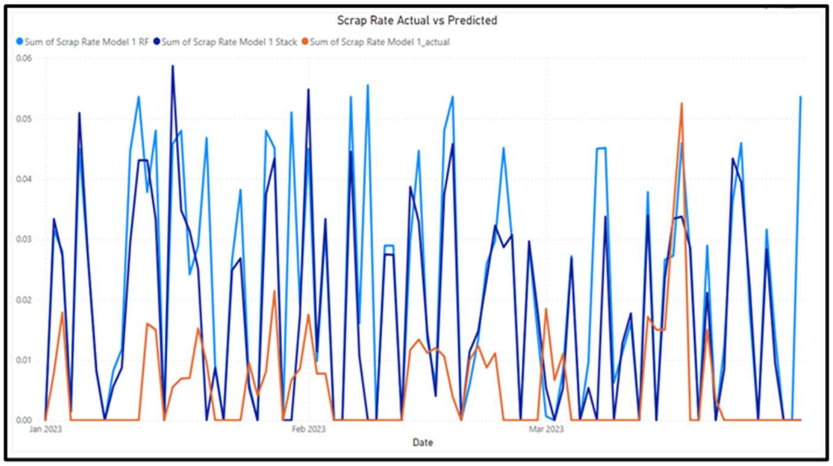


Figure 14. Actual VS Predicted Model 1

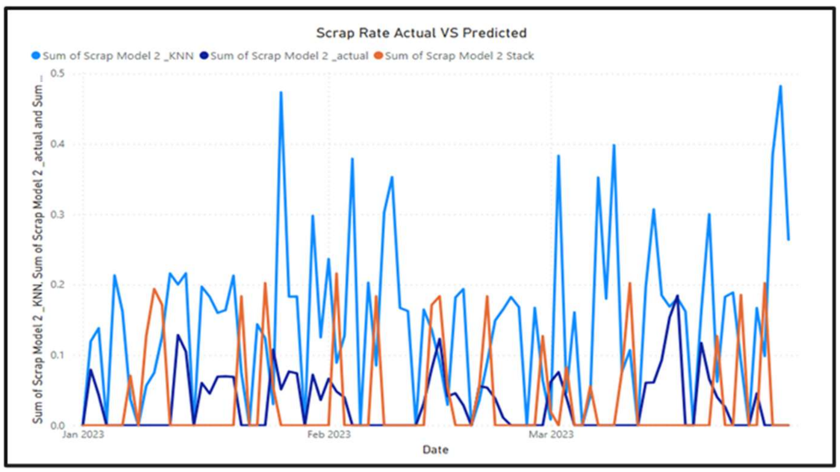


Figure 15. Actual VS Predicted Model 2

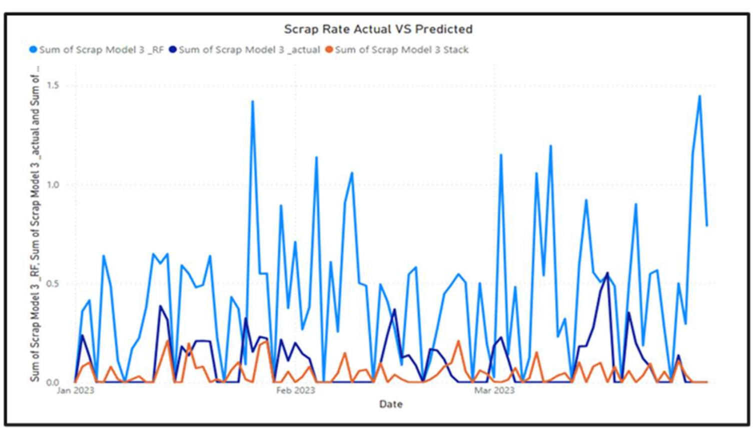


Figure 16. Actual vs Predicted Model 3

As shown in Figure 17, 18, and 19, the results indicate that the projected scrap rates for Models 1, 2, and 3 exhibit some irregularity. This variability is partly due to excluding Sunday data, as the factory is typically shut down on that day. Despite this introduced irregularity, the figure achieves the objective of predicting product scrap rates. The mean value for all models is below 0.08, demonstrating both high accuracy and precision. Such reliable forecasts can significantly support preventive maintenance efforts by helping prevent equipment anomalies or unexpected increases in scrap rates.

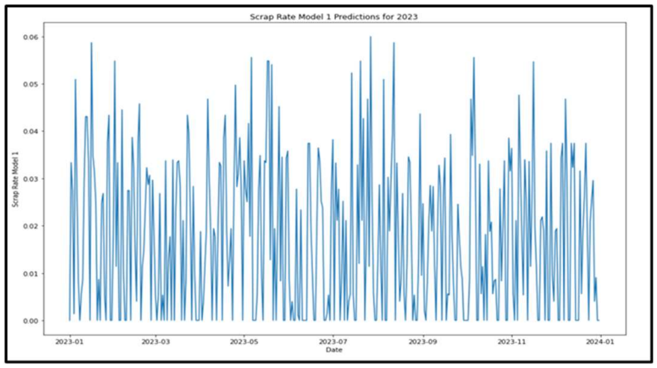


Figure 17. Result Predicted Model 1

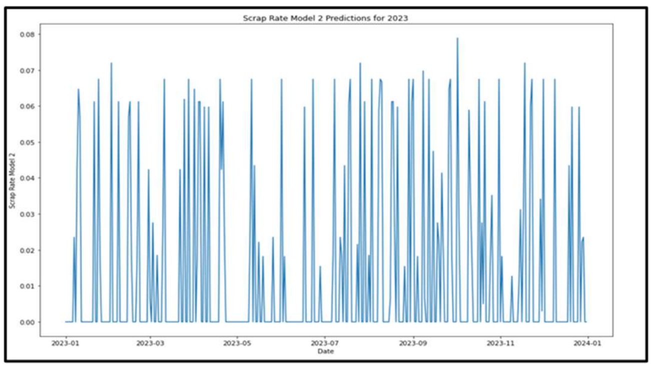


Figure 18. Result Predicted Model 2

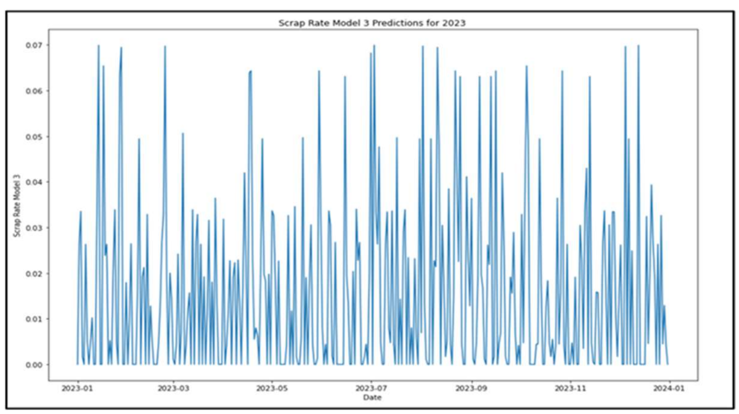


Figure 19. Result Predicted Model 3

Percentage Different Error Between Actual and Predicted Data

Table 3 indicates that Model 1's Mean Absolute Percentage Error (MAPE) was 75.64%, suggesting its forecasts were off by an average of about three-quarters of the actual value. Model 2 performed somewhat better with a MAPE of 42.58%, but this still reflects a considerable deviation. Model 3 yielded the most accurate forecasts, recording a MAPE of 39.35%. While this error rate is still high, it is notably lower than the other two models.

In summary, Model 3 (HM73-10C300LFTR) was the most accurate of the three according to MAPE results. Nonetheless, a 39.35% error rate implies there is room for improvement. Future work could involve refining model parameters, experimenting with alternative algorithms, or incorporating additional variables to enhance prediction accuracy.

Table 3. MAPE result using the Best Model

Keyword	Model	MAPE (%)
Model 1	HA00-08464LFTR	75.64
Model 2	HA00-17359ALFTR	42.58
Model 3	HM73-10C300LFTR	39.35

Table 4. Data Actual and Predicted

Date	Model 1		Model 2		Model 3	
	Actual	Predicted	Actual	Predicted	Actual	Predicted
1/1/2023	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2/1/2023	0.007857	0.039392	0.027500	0.026410	0.043750	0.056516
3/1/2023	0.017833	0.027770	0.015692	0.030988	0.010000	0.072100
4/1/2023	0.000000	0.000000	0.000000	0.001845	0.000000	0.000055
5/1/2023	0.000000	0.000000	0.000000	0.050570	0.000000	0.000000
6/1/2023	0.000000	0.000000	0.000000	0.024763	0.000000	0.000000
7/1/2023	0.000000	0.000000	0.000000	0.005484	0.000000	0.001775
8/1/2023	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9/1/2023	0.000000	0.000000	0.000000	0.005484	0.000000	0.018987
10/1/2023	0.000000	0.000000	0.000000	0.013252	0.000000	0.056276
11/1/2023	0.000000	0.000000	0.000000	0.070256	0.000000	0.055354
12/1/2023	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13/1/2023	0.016000	0.051100	0.060000	0.034462	0.052500	0.000000
14/1/2023	0.015000	0.036367	0.055000	0.064255	0.035000	0.000000
15/1/2023	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16/1/2023	0.005500	0.091671	0.024857	0.053000	0.030000	0.000000
17/1/2023	0.006889	0.034749	0.021250	0.064390	0.017143	0.000000
18/1/2023	0.007000	0.024515	0.031667	0.024337	0.030667	0.000000
19/1/2023	0.015231	0.024515	0.037000	0.023981	0.017625	0.000000
20/1/2023	0.009545	0.000000	0.023750	0.074887	0.035667	0.000000
21/1/2023	0.000000	0.009162	0.000000	0.005484	0.000000	0.051889
22/1/2023	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
23/1/2023	0.000000	0.027770	0.000000	0.020673	0.000000	0.091396
24/1/2023	0.000000	0.032957	0.000000	0.019132	0.000000	0.059099
25/1/2023	0.009545	0.003144	0.052500	0.005484	0.045714	0.003245
26/1/2023	0.004040	0.000000	0.022444	0.045769	0.025000	0.000000
27/1/2023	0.008077	0.031727	0.024800	0.062600	0.043750	0.000000
28/1/2023	0.021400	0.031495	0.024667	0.057968	0.028000	0.000000
29/1/2023	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
30/1/2023	0.006679	0.000000	0.025250	0.074887	0.040000	0.000000
31/1/2023	0.008560	0.014994	0.027778	0.021285	0.000000	0.090816

Dashboard for Monitoring the Data

The dashboard, built in Power BI, transforms the Jupyter Notebook analysis into a visually engaging interface with filtering capabilities, offering fresh insights into the data. As discussed in the method section, the u-chart is essential for monitoring process stability. By integrating the u-chart into the dashboard, users can easily correlate process control information with other relevant datasets.

Figures 20, 21, and 22 illustrate the Power BI dashboard. Card visualizations provide key metrics for all data, while filtering options allow users to focus on specific models or machines as needed. The line charts in Figures 21 and 22 utilize a basic control chart, including upper, center, and lower limits. These limits apply to the entire dataset, not just a single model. When filters are applied, the scrap rate is generally under control, though occasional anomalies still appear on certain dates.

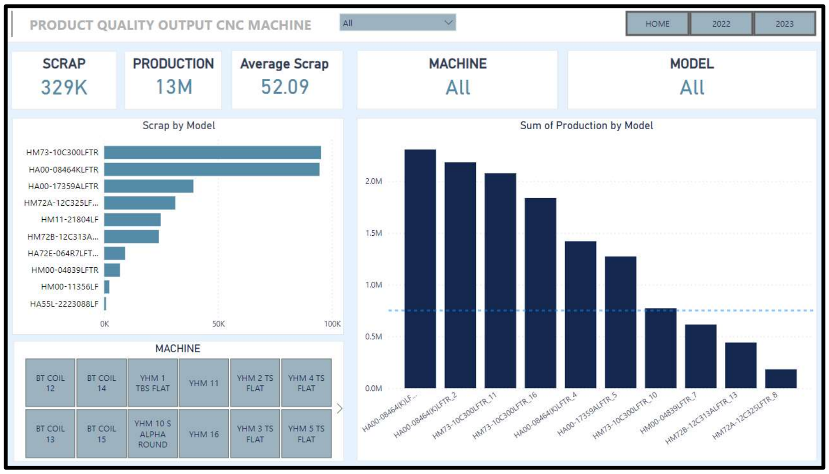


Figure 20. Dashboard HOME

Figure 21. Dashboard 2022

Figure 22. Dashboard 2023

4.2. Discussion

Air coil output quality measurement can be obtained through the analysis of scrap data generated by the CNC machine. At various instances, the occurrence of diverse air coil scrap is observed. These varied scrap occurrences indicate a complex interplay of factors, including tool wear, machine calibration, material inconsistencies, or even operator skill levels. Each factor potentially influences the final product quality (references).

Visualizing the scrap rates and production data via the dashboard provides a valuable lens through which anomalies and fluctuations become immediately apparent. The *u*-chart, as a monitoring instrument, enables practitioners to detect variations beyond normal statistical limits (Ieren et al., 2020; Tang et al., 2019). Such an analysis highlights the importance of establishing evidence-based benchmarks for identifying when the process deviates from its intended parameters. However, while *u*-charts help reveal non-conformance, they do not inherently explain the root causes, thus requiring deeper diagnostic efforts—like root cause analysis or additional process measurements—to fully understand why anomalies persist.

When forecasting scrap rates, the decision to employ advanced machine learning models—KNN and Random Forest—reflects an acknowledgment that simple statistical tools may be insufficient to capture the complexity of manufacturing data (Soori, et al., 2023; Zhang & Jiang, 2019). Nevertheless, the models' initial performance fell short of expectations, likely due to factors such as data quality, unmodeled process variability, or insufficient feature engineering. In response, the study adopted a stacking approach, combining multiple predictive models to harness their individual strengths. Although stacking notably improved performance for Models 2 and 3, Model 1 did not exhibit the same gains. This discrepancy may point to unique process conditions, variations in machine health, or parameter interactions not captured by the ensemble.

Choosing the stacking method over RF alone, even for Model 1, where RF showed higher peaks, was guided by stacking's closer alignment with actual data. This choice highlights a key challenge in predictive modeling: finding the right balance between pure predictive accuracy and stable, reliable performance. Opting for the model that best matches real-world patterns, rather than one that is only statistically strong, demonstrates a practical approach to model selection.

Concluding that ensemble stacking is the preferred method for predicting scrap rates represents a move toward more advanced, data-driven preventive maintenance strategies. It suggests that no single model type will always perform best and that combining various approaches can yield more robust result. However, it is important to recognize that forecasting accuracy can be influenced by factors such as production schedule changes, maintenance timing, material quality, and evolving customer requirements.

While this study meets its primary goal of improving predictive capabilities, it also suggests the need for ongoing refinement. Future research may incorporate additional data sources—like sensor readings, operator logs, or environmental conditions—to improve the models' contextual understanding. Further enhancements could involve fine-tuning model parameters, revisiting the chosen features, or experimenting with more advanced algorithms like gradient boosting or neural networks. In this way, these findings both confirm the value of ensemble methods in preventive maintenance and encourage further investigation into the complex factors affecting CNC machine performance and product quality.

5. Conclusion

The study reveals the prospective application of *u*-charts and machine learning regression models to enhance preventive maintenance strategies for CNC machines. By analyzing production and scrap data, we effectively predicted scrap rates, enabling more informed decision-making and resource allocation.

The *u*-charts proved valuable for visualizing fluctuations in scrap rates over time, making it easier to identify patterns and anomalies. Understanding these patterns led to a significant reduction in waste, which highlights the importance of data visualization in production management.

Furthermore, the efficacy of regression models based on machine learning by using random forest (RF) and *k*-Nearest Neighbor (kNN) showed strong predictive capabilities. Training these models with historical production data allowed us to anticipate scrap rates accurately, reduce downtime, and improve overall efficiency.

In addition, our study has revealed the complex dynamics relationship between u-charts and machine learning models. *U*-charts helped identify areas needing attention, guiding model refinement by identifying key variables and unusual data points. In turn, machine learning forecasts provided insights that could be monitored and tracked through u-charts, further improving process oversight.

This study's findings are constrained by several limitations. The results depend on high-quality, comprehensive data and may not generalize to other machines or contexts. Statistical assumptions about distributions and independence may not hold, and model selection was limited. Practical factors like cost, scheduling, and training were not fully addressed.

Future research should explore more advanced machine learning methods, apply the approach in varied settings, incorporate real-time analytics, and consider broader factors such as economic constraints and operator skills. This would enhance reliability, applicability, and overall understanding of predictive maintenance strategies.

To summarize, the integration of these methodologies into the preventive maintenance plan of CNC machines has demonstrated advantages in optimizing manufacturing activities, enhancing the standard of the final product, and minimizing material waste. The present study highlights the prospects of utilizing data analysis and forecasting tools to transform the maintenance tactics implemented in the manufacturing sector.

Acknowledgment

I would like to express my gratitude to my senior employees, Encik Faizul, Encik Zul, Puan Hajar, and Puan Hayati who assisted and encouraged me all the way through the process of finishing this project as well as all the years that I spent studying

Declaration of Conflict of Interest

The authors declare no potential conflicts of interest related to the research, writing, and/or publication of this article

References

Alquthami, T., Zulfiqar, M., Kamran, M., Milyani, A. H., & Rasheed, M. B. (2022). A performance comparison of machine learning algorithms for load forecasting in smart grid. *IEEE Access*, *10*, 48419–48433. https://doi.org/10.1109/ACCESS.2022.3171270

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Dewi, S., Alhilman, J., & Atmaji, F. T. D. (2020, May). Evaluation of effectiveness and cost of machine losses using overall equipment effectiveness (OEE) and overall equipment cost loss (OECL) methods, a case study on Toshiba CNC machine. In *IOP Conference Series: Materials Science and Engineering* (Vol. 847, No. 1, p. 012020). Institute of Physics Publishing. https://doi.org/10.1088/1757-899X/847/1/012020

Haddad, T., Shaheen, B. W., & Németh, I. (2021). Improving overall equipment effectiveness (OEE) of extrusion machine using lean manufacturing approach. *Manufacturing Technology*, *21*(1), 56–64. https://doi.org/10.21062/mft.2021.006

Ieren, T. G., Kuje, S., Asongo, A. I., & Eraikhuemen, I. B. (2020). Application of statistical quality control in monitoring the production, packaging and marketing process of sachet water. *Journal of Scientific Research and Reports*, *26*(9), 32–45. https://doi.org/10.9734/jsrr/2020/v26i930307

Lee, H., Kang, D. H., & Jeong, S. C. (2022). A study on the predictive maintenance system for CNC facilities based on intelligent cutting tool load prediction. *ICIC Express Letters, 16*(3), 265–273. https://doi.org/10.24507/icicel.16.03.265

Liu, W., Zhang, S., Lin, J., Xia, Y., Wang, J., & Sun, Y. (2022, August 1). Advancements in accuracy decline mechanisms and accuracy retention approaches of CNC machine tools: A review. *International Journal of Advanced Manufacturing Technology, 121*(11–12), 7087–7115. https://doi.org/10.1007/s00170-022-09720-0

Mishra, R., Kiran, M. S. N. S., Maheswaram, M., Upadhyay, A., & Singh, B. (2023, June). Investigation of optimal feature for milling chatter identification using supervised machine learning techniques. *Journal of Engineering Research*, 100138. https://doi.org/10.1016/j.jer.2023.100138

Moghaddam, K. S. (2015, February). Preventive maintenance and replacement optimization on CNC machine using multiobjective evolutionary algorithms. *International Journal of Advanced Manufacturing Technology, 76*(9–12), 2131–2146. https://doi.org/10.1007/s00170-014-6391-5

Nagahisarchoghaei, M., Dodd, J., Nagahi, M., Ghanbari, G., & Poudyal, S. (2020, October). Analysis of a warranty-based quality management system in the construction industry. In *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI 2020)*. IEEE. https://doi.org/10.1109/ICDABI51230.2020.9325692

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015, August). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews, 71*, 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001

Ruiz Rodríguez, M. L., Kubler, S., de Giorgio, A., Cordy, M., Robert, J., & Le Traon, Y. (2022, December). Multi-agent deep reinforcement learning based predictive maintenance on parallel machines. *Robotics and Computer-Integrated Manufacturing, 78*, 102406. https://doi.org/10.1016/j.rcim.2022.102406

Shahin, M., Chen, F.F., Hosseinzadeh, A. *et al.* (2023). Using machine learning and deep learning algorithms for downtime minimization in manufacturing systems: an early failure detection diagnostic service. *International Journal of Advanced Manufacturing Technology* 128, 3857–3883. https://doi.org/10.1007/s00170-023-12020-w

Singh, S., Khamba, J. S., & Singh, D. (2021, April 1). Analysis and directions of OEE and its integration with different strategic tools. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, 235*(2), 594–605. https://doi.org/10.1177/0954408920952624

Soori, M., Arezoo, B., & Dastres, R. (2023, January). Machine learning and artificial intelligence in CNC machine tools: A review. *Sustainable Manufacturing and Service Economics*, 100009. https://doi.org/10.1016/j.smse.2023.100009

Tang, D., Tanli, M., Jiang, Y., Wan, X., & Peng, R. (2019, September). Product quality monitoring of Shewhart chart based on function integration for manufacturing factory. In *Journal of Physics: Conference Series, 1302*(4), 042044. Institute of Physics Publishing. https://doi.org/10.1088/1742-6596/1302/4/042044

Traini, E., Bruno, G., D'Antonio, G., & Lombardi, F. (2019). Machine learning framework for predictive maintenance in milling. *IFAC-PapersOnLine, 52*(13), 177–182. https://doi.org/10.1016/j.ifacol.2019.11.172

Tsarouhas, P. H. (2019, January). Improving operation of the croissant production line through overall equipment effectiveness (OEE): A case study. *International Journal of Productivity and Performance Management*, *68*(1), 88–108. https://doi.org/10.1108/IJPPM-02-2018-0060

Tsarouhas, P. H. (2020, May). Overall equipment effectiveness (OEE) evaluation for an automated ice cream production line: A case study. *International Journal of Productivity and Performance Management, 69*(5), 1009–1032. https://doi.org/10.1108/IJPPM-03-2019-0126

Zhang, C., & Jiang, P. (2019). Sustainability Evaluation of Process Planning for Single CNC Machine Tool under the Consideration of Energy-Efficient Control Strategies Using Random Forests. *Sustainability*, *11*(11), 3060. https://doi.org/10.3390/su11113060