



Vaasan yliopisto
UNIVERSITY OF VAASA

Syed Mahmood ul Hassan

Predictive Maintenance in IoT Using NLP Techniques

Master Thesis

School of Technology and Inno-
vations
Master of Sustainable and Au-
tonomous Systems

Vaasa 2025

UNIVERSITY OF VAASA**School of Technology and Innovations**

Author:	Syed Mahmood ul Hassan		
Title of the thesis:	Predictive Maintenance in IoT Using NLP Techniques		
Degree:	Master of Sustainable and Autonomous Systems		
Supervisor:	Mohammed Elmusrati		
Co-supervisor	Elham Ahmadi		
Year:	2025	Pages:	82

ABSTRACT:

In this thesis we study how NLP techniques can be integrated with Internet of Things (IoT) sensor data, to be used for predictive maintenance in industrial environments. The main objective was to build models that can predict whether an equipment will fail in the next 7 days using the `iot_predictive_maintenance` dataset. Sensor readings include temperature, vibration, humidity, and pressure etc. along with unstructured textual maintenance logs. Two sets of the feature were built: one containing only numeric sensor data and the other with features from TextBlob, namely TF-IDF vectors and polarity scores on sentiment, to create a feature set. Logistic Regression, Random Forest, Gradient Boosting, and XGBoost models were evaluated multiple times. It was found that models incorporating both sensor and NLP features significantly outperformed those based solely on sensor data. Among the evaluated models, the best performance was achieved by the XGBoost model using the combined feature set, which attained an accuracy of 0.973, an F1 score of 0.889, and a ROC AUC of 0.993. These results confirm that the textual information in maintenance logs contains important failure aspects not expressed from numerical data alone. This work demonstrates practicality of fusion models in predictive maintenance, enabling a scalable and robust solution for smart manufacturing.

KEYWORDS: Predictive Maintenance, IoT, Natural Language Process (NLP), Machine Learning (ML), Fault Prediction

Contents

1	Introduction	9
1.1	Background and motivation	12
1.2	Problem statement	13
1.3	Research objectives	14
1.4	Research questions	15
1.5	Methodology overview	15
1.5.1	Data preprocessing	16
1.5.2	Feature engineering	17
1.5.3	Model Development and Training	18
1.5.4	Evaluation metrics	19
1.6	Significance and contribution	20
2	Literature Review	21
2.1	Traditional predictive maintenance and sensor-based approaches	22
2.2	Unstructured data and the case for NLP in maintenance	24
2.3	Hybrid approaches: Fusing text with sensor data	26
2.4	Advances in NLP models for industrial applications	27
2.5	Evaluation practices and real-world validation	29
2.6	Ethical, practical, and operational considerations	30
2.6.1	Human-in-the-Loop (HITL) Approaches	30
2.6.2	Privacy and Data Governance	31
2.6.3	Transparency and Accountability	31
2.6.4	Operational Challenges	31
2.7	Summary of Gaps and Opportunities	32
3	Methodology	35
3.1	Dataset overview	35
3.1.1	Sensor data	36
3.1.2	Maintenance logs	37
3.1.3	Objectives	37
3.1.4	Dataset limitations	38

3.1.5	Relevance to industry	39
3.2	Data pre-processing	39
3.2.1	Sensor data	39
3.2.2	Maintenance log	40
3.3	Feature engineering	41
3.4	Model development	41
3.5	Model evaluation metrics	42
3.6	Findings of preliminary methodology	43
4	Case study – Predictive Maintenance using Sensor and Maintenance Log Data with NLP	45
4.1	Dataset overview	45
4.2	Data pre-processing	47
4.2.1	Sensor data processing	47
4.2.2	Maintenance log processing	47
4.3	Model development	48
4.3.1	Feature set	48
4.3.2	Model selection	49
4.3.3	Model evaluation	49
4.4	Summary	50
4.5	Insights and implications	51
5	Results and discussion	53
5.1	Model performance	53
5.1.1	Logistic regression	57
5.1.2	Random forest	58
5.1.3	Gradient boosting	59
5.1.4	XGBoost	61
5.2	Best model analysis: XGBoost with NLP-Enhanced Features	64
5.3	Best Performing Model: XGBoost (NLP-Enhanced)	65
5.4	Discussion	66
5.4.1	Impact of feature set	67

5.4.2	Model comparison	67
5.4.3	Confusion matrix insights	68
5.4.4	Practical implications	69
6	Conclusion	70
6.1	Summary of work	70
6.2	Key findings	71
6.3	Critical evaluation	72
6.4	Practical implications	73
6.5	Limitations and future directions	73
	References	75
	Appendices	Error! Bookmark not defined.
	Appendix 1. Appendix title	Error! Bookmark not defined.
	Appendix 2. Another appendix title	Error! Bookmark not defined.

Figures

Figure 1: Dataset of IoT devices	10
Figure 2: Model architecture	19
Figure 3: Dataset pre-processing	36
Figure 4: Failure rate per device	38
Figure 5: Sensor readings and failure rate of devices	41
Figure 6: Confusion matrix of sensor only XGBoost	50
Figure 7: XGBoost confusion matrix of sensor and textual combined feature	51
Figure 8: Model comparison w.r.t accuracy	54
Figure 9: Logistic regression sensor only model	57
Figure 10: Logistic regression NLP enhanced model	58
Figure 11: Random forest sensor feature set	58
Figure 12: Random forest NLP enhanced model	59
Figure 13: Gradient boosting sensor only model.	60
Figure 14: Gradient Boosting with NLP Features	60
Figure 15: XG Boost Sensor Model Summary	61
Figure 16: XG Boost NLP Model Summary	62
Figure 17: ROC Curve of XG Boost	62
Figure 18: Top 10 feature importance	67
Figure 19: Model comparison w.r.t to F1 score	68
Figure 20: Model comparison summary.	68
Figure 21: Confusion matrix of best performing XGBoost model.	69

Tables

Table 1: IoT sensors and their functions	16
Table 2: Raw sensor features	18
Table 3: Sample TF-IDF terms from maintenance logs	40
Table 4: Train-Test split overview	49
Table 5: Comparative ROC AUC Scores	55

Table 6: Overall comparion of models	56
Table 7: Sensor only feature set.	63
Table 8: NLP enhanced feature set.	63
Table 9: Confusion matrix textual representation.	64
Table 10: Sentiment polariy distribution in logs	64

Abbreviations

AI	Artificial Intelligence
AUC	Area Under Curve
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
BERT	Bidirectional Encoder Representations from Transformers
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GBM	Gradient Boosting Machine
GDPR	General Data Protection Regulation
HITL	Human-in-the-Loop
IIoT	Industrial Internet of Things
IoT	Internet of Things
ML	Machine Learning
MPM	Modern Predictive Maintenance
NER	Named Entity Recognition
NLP	Natural Language Processing
PdM	Predictive Maintenance
ROC	Receiver Operating Characteristic
RUL	Remaining Useful Life
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Over-sampling Technique
TF-IDF	Term Frequency–Inverse Document Frequency
TPR	True Positive Rate

XGBoost

eXtreme Gradient Boosting

1 Introduction

For example, with Industry 4.0, the era of digital transformation, cyber-physical systems, and automation, as well as smart sensing technologies are integrated into industrial processes. Within this revolution sits PdM, an advanced maintenance strategy that is poised to predict the failures of equipment ahead of their occurrences, thus preventing unscheduled downtimes, improving operational efficiency, and helping extend the operating lifespan of critical assets that are deemed indispensable. With the IoT having greatly boosted this capability, advances in sensor embedded systems allow for continuous monitoring of machines and generation of vast amounts of structured, time series data from parameters like temperature, vibration, pressure, and humidity.

Sensor driven predictive models are remarkably successful yet in most current frameworks, the unstructured textual data generated along sensors is underutilized to date. Examples include maintenance logs, fault descriptions, incident reports, and technician observations that usually contain rich contextual information that is hard to capture by numerical data only. According to (Kang et al., 2020), such exclusion could fail because of incomplete fault detection, inefficient maintenance scheduling, and delayed early warning of failure, leaving the promise of PdM systems unrealized.

To address this, this research conducts a study to build a new machine augmented PdM model, one that will complement both the structured IoT sensor data and unstructured textual records. Specifically, this study presents a dual layer predictive framework where NLP techniques are used to extract and encode important features in the maintenance logs. Finally, these sensors are combined with TF-IDF and sentiment analysis obtained with TextBlob to train advanced ML models that can better predict failure.

The work here studies PdM using separated data from a variety of IIoT purpose-built sensors. For collecting real time measurement of temperature, pressure, flow rate and environmental conditions, ThermoTrack-B3, SensorHub-A1, FlowMeter-E9 and EnviroMon-D2 devices came out to be very important. Based on maintenance logs and

sensor readings, the predictive models that were developed in turn rely on these sensor readings to predict equipment failures before they take place.

In this study, our primary dataset shown in figure 1, which was used for doing such research was from Kaggle, it has sensor reading of the equipment real time and corresponding logs of maintenance as shown in figure 1. The structured part included key operational metrics such as temperature, humidity, vibration, pressure, while unstructured part was technician entered head provider symptoms and fau definitions and anomalies observed. This research applies linguistic intelligence to an integration of these two modalities of data, addressing one of the central problems in PdM: contextual understanding.

device_name	temperature	vibration	humidity	pressure	maintenance	failure_within_7_days
ThermoTri	79.96714	3.199678	54.42639	104.1656	Minor weat	0
SensorHub	73.61736	2.962317	33.96059	92.37067	Sensor ma	1
FlowMeter	81.47689	2.529815	50.92923	99.50185	Dust accur	0
EnviroMor	90.2303	2.176532	36.65980	102.8055	No issues c	0
EnviroMor	72.65847	2.849112	78.49411	102.2192	No issues c	0
VibeSense	72.65863	2.696743	86.86128	114.7652	Pressure r	0
ThermoTri	90.79213	2.947597	34.33899	103.049	Sensor ma	1
ThermoTri	82.67435	2.817586	87.30691	96.27973	Unusual vi	0
HumidityG	70.30526	3.024776	61.3546	100.8227	Dust accur	0
SensorHub	80.4256	2.232382	47.97394	92.41877	Sensor ma	1
SensorHub	70.36582	3.158697	34.61172	100.897	Minor weat	0
ThermoTri	70.3427	2.5988	60.03746	97.13472	Routine ch	0
EnviroMor	77.41962	3.53763	77.67093	105.877	Sensor ma	1
EnviroMor	55.8672	2.155406	72.42519	98.5523	Dust accur	0
SensorHub	57.75082	3.367982	33.01356	100.7162	Sensor ma	1
EnviroMor	69.37712	2.598955	34.37411	98.12221	Sensor ma	0
HumidityG	64.87169	2.174291	54.1724	109.9943	Pressure r	0
EnviroMor	78.14247	2.258057	47.71743	99.69379	Unusual vi	0
TempCont	65.91976	2.339826	43.94306	110.4628	Dust accur	0
FlowMeter	60.87696	2.712083	46.86027	105.3708	No issues c	0
SensorHub	89.65649	2.761418	78.20896	103.7103	High temp	0
VibeSense	72.74224	2.21315	85.75368	103.1437	Unusual vi	0
HumidityG	75.67528	2.487823	54.30616	103.269	Unusual vi	0
PressurePr	60.75252	3.571135	84.36666	91.66164	Routine ch	1
FlowMeter	69.55617	3.363772	49.28974	99.90558	High temp	0
VibeSense	76.10923	2.718162	58.58622	105.5258	Unusual vi	0
EnviroMor	63.49006	2.519002	43.56174	101.0555	Minor weat	0

Figure 1: Dataset of IoT devices

In this methodology, the textual data are pre-processed namely cleaning it of noise, standardizing it to the language, tokenizing content and then feature extracted through TF-IDF to focus on the importance of the words based on the documents as well as to run sentiment analysis indicating whether the logs were having polarity or not showing signs of distress or failure. Based on these, two unique datasets were formed: first, a

dataset comprised solely of sensor data, and second, a combined dataset including NLP derived features.

These approaches were evaluated with for four ML models: Logistic Regression, Random Forest, Gradient Boosting and an optimized XGBoost classifier. The key performance metrics for these models were accuracy, precision, recall and F1-score along with ROC-AUC. Results showed that sensor data-based models provided significantly worse results than those enhanced by the NLP dataset. We found the most notable improvements of the XGBoost model, which reached a remarkable accuracy of 0.973, an F1-score of 0.889 and a ROC-AUC of 0.993, therefore proving its stability and precision, it is in identifying possible failures.

Additionally, confusion matrix analysis revealed that NLP integrated models generated less FPs and FNs than the models without semantic insights from textual logs, which means that semantic insight from textual logs imbedded in the models leads to a more robust differentiator between failure and non-failure equipment. Finally, the sentiment polarity scores provided value by providing information as to the technician's tone and urgency, which correlate with real failure events.

Thus, this thesis provides additional contributions to the developing knowledge base about intelligent maintenance systems by demonstrating that NLP can be a very powerful complement to sensor-based models for PdM. The results show the need for multi-modal data fusion in industrial analytics and a scalable framework that can be applied in the general industrial IoT eco-systems. The proposed hybrid PdM model captures both quantitative and qualitative signals, making it possible to develop more resilient and is better and less expensive maintenance solutions in smart manufacturing environments.

In the following chapters, the thesis will reveal the methodology and experimental setup details, include details of experimental results and analysis, as well as conclusions, limitations, and future research.

1.1 Background and motivation

Thanks to the exponential growth and the widespread distribution of IoT devices, industrial operations can now be continuously and in time, monitor as granular data as possible in the manufacturing field, oil and gas, aerospace, and transportation (Compare et al., 2020). By sensing high resolution structured data streams of critical machine parameters of temperature fluctuation, vibration frequencies, pressure levels, humidity change, and operational cycle, these are smart, interconnected sensors. PdM systems can analyse these data points, identify anomalies, predict future faults, and advise on maintenance schedules, which would help increase system reliability and operational efficiency in general.

However, while these advancements enable improved detection of failures, it is still limited with the semantic depth missing to truly grasp the root causes of failure and contextual factors of maintenance issues. If sensor-based data is used, the models are typically optimized with quantitative metrics at the expense of consideration of qualitative indicators that are also important for interpretation. For example, there is information provided by technicians in comments, root cause descriptions, work arounds notes, and previous fault narratives in maintenance logs, service reports, or inspection notes; information that is not given in the form of numerical sensors. Often these unstructured text entries relate to such anomalies in natural language that describe recurrent issues and provide contextual clues as well as metadata values relating to environmental or operational conditions that may not otherwise be monitored.

According to Usuga-Cadavid et al (2022), the use of such textual records in PdM frameworks can unveil unobserved latent patterns and previously uncorrelated failure modes with machine behaviour (P. U. Cadavid et al., 2021). It enriches the predictive modelling procedure, providing a frequent and interpretable mechanism to analyse equipment condition. Although these unstructured data sources have potential, they are not yet

used in mainstream industrial analytics due to challenges in processing such unstructured data which include variation in language, domain jargon specific to a problem, misspellings, and variations in formatting.

Hence, the reason for this research is to bridge the gap between the structured analytic constructor stemming from sensor analytics and the unstructured textual intelligence, applying the NLP approach which transforms free text record into machine readable features. The study intends to integrate numerical sensor data with NLP derived features namely, TF-IDF vectors and sentiment polarity scores into building more robust holistic PdM model. The contribution is in the use of this interdisciplinary approach that integrates the strengths of data driven engineering and computational linguistics to provide for smarter, context aware maintenance solution that can yield deeper operational insight and more accurate fault predictions.

1.2 Problem statement

In modern industrial environments, PdM has become an essential strategy for minimizing unplanned equipment downtime, the reduction of maintenance costs and increase overall system reliability. In fact, they have led to the advent of sophisticated PdM models based not on one or two data points at a time but rather on continuous streams of numerical data (vibration, temperature, pressure, etc.) to detect anomalies and predict faults. However, to date unstructured textual information available in these models is underutilized.

Maintenance logs, technician notes, incident reports and service documentation are generated routinely in large volumes daily during industrial operations. In fact, often these records have a lot of qualitative insights, including early warnings signs, context about previous faults, human observation, and domain specific knowledge, which cannot be captured by sensors alone. Unfortunately, conventional PdM frameworks again

ignore these unstructured data sources by focusing on only structured sensor inputs. This therefore means that the current models tend to be less aware of the context, which then constrains their ability in being able to predict correctly or explaining the reason of why there is an anomaly or equipment failure.

Current practice leaves this gap, increasing the chances for unexpected equipment breakdowns, which cause poor productivity, safety, and operational continuity, and with it, several pressing challenges of incomplete or inaccurate fault detection, suboptimal maintenance scheduling, arising. Taking this opportunity to exploit textual data under exploited, would yield rich human centric insights that complement predictive analytics. Therefore, the main problem this research addresses is to design, develop and validate an NLP empowered PdM framework to combine quantitative sensors data with qualitative textual data natively. This model unites the benefits of NLP techniques with those of structured data sources, by incorporating features extracted through Forward NLP, namely TF-IDF vectorization and sentiment analysis. The goal is to increase the prediction accuracy, interactivity, and reliability of PdM systems and to approach the failure prediction, anticipating failures, and developing optimal industrial maintenance strategies.

1.3 Research objectives

To address the outlined problem, the thesis sets forth the following objectives:

1. To analyze an IoT sensor dataset and identify relevant unstructured textual features.
2. To preprocess and clean maintenance logs for NLP application, including tokenization, removal of noise, and standardization.
3. To apply TF-IDF and sentiment analysis to extract meaningful features from the logs.

4. To integrate textual features with sensor data to develop ensemble-based ML models for failure prediction.
5. To compare the performance of NLP-enhanced models with traditional sensor-only models using evaluation metrics such as accuracy, F1 score, and AUC.

1.4 Research questions

The study aims to answer the following research questions:

- What kind of improvements in predictive outcome can be made when combining NLP with IoT sensor data?
- What preprocessing and feature engineering steps are essential for converting unstructured maintenance logs into actionable predictive inputs?
- How do models that include NLP-derived features compare in performance to those that utilize only structured sensor data?

1.5 Methodology overview

The methodology employed in this research followed a systematic, multi-stage pipeline designed to develop, integrate, and evaluate a hybrid PdM model that combines structured IoT sensor data with unstructured maintenance log entries. The process was broadly divided into four key stages: data preprocessing, feature engineering, model development, and performance evaluation. The primary dataset used for experimentation was the publicly available IoT PdM dataset, which included time-stamped sensor measurements—such as temperature, humidity, pressure, and vibration—alongside free-text maintenance logs describing fault occurrences and technician observations.

1.5.1 Data preprocessing

The preprocessing phase was critical in preparing both structured and unstructured data for downstream analysis. The textual component of the dataset (maintenance logs) underwent a series of standard NLP preprocessing steps, which included:

- Lowercasing all text to ensure uniformity.

$$\log_i \text{ clean} = \text{Clean}(\log i)$$

- Removal of punctuation, digits, and special characters to eliminate irrelevant noise.
- Tokenization and stop-word removal to retain only meaningful words.
- Lemmatization to reduce words to their root forms for semantic consistency.

Table 1 lists the IoT sensors used in the study, detailing their types and specific measurement focuses. These devices collectively capture a wide range of environmental and system parameters, including temperature, humidity, pressure, flow, and vibration.

Table 1: IoT sensors and their functions

Device name	Type	Measurement focus	Description
ThermoTrack-B3	Temperature sensor	Temperature	Tracks thermal conditions
SensorHub-A1	Multi-sensor hub	Combined set of sensors	Aggregate data from multiple sensor types
FlowMeter-E9	Flow Sensor	Fluid/Gas Flow	Monitors flow rate in pipelines
EnviroMon-D2	Environmental Monitor	Temperature, Humidity	Measures ambient environmental parameters

VibeSense-C7	Vibration Sensor	Vibration	Detects mechanical vibrations in equipment
HumidityGuard-G8	Humidity Sensor	Humidity	Monitors moisture in the environment
TempControl-H6	Temperature Controller	Temperature Regulation	Controls and adjusts system temperature
PressurePro-F4	Pressure Sensor	Pressure	Measures internal system pressure

Following this, the processed text was transformed into numerical features using TF-IDF vectorization. A feature cap of 50 terms was selected based on exploratory analysis and dimensionality constraints to capture the most informative tokens across the corpus. Additionally, sentiment polarity scores were extracted using the TextBlob library, providing a scalar measure of the emotional tone of technician comments (ranging from negative to positive sentiment). Then we use these scores to capture implicit cues in human written log that might represent urgency or seriousness of these issues.

1.5.2 Feature engineering

To explore the added value of integrating unstructured data, two distinct feature sets were constructed:

- **Sensor-Only Features:** This included the original numerical attributes from the IoT sensor data (temperature, humidity, pressure, and vibration), representing the conventional approach to PdM

$$x_i^{sensor} = [t_i, v_i, h_i, p_i] \in R^4$$

- **NLP-Augmented Features:** This composite feature set combined the sensor data with the TF-IDF text vectors and sentiment polarity scores. The goal was to enrich the numerical representation with linguistic cues and contextual signals embedded in maintenance logs.

$$x_i^{nlp} = [t_i, v_i, h_i, p_i, s_i, w_i] \in R^{M+5}$$

Table 2: Raw sensor features

Feature name	Source device	Unit	Data type
Temperature	ThermoTrack-B3	°C	Numeric
Vibration	VibeSense-C7	mm/s	Numeric
Humidity	HumidityGuard-G8	%RH	Numeric
Pressure	PressurePro-F4	kPa	Numeric
Flow Rate	FlowMeter-E9	L/min	Numeric

Table 2 presents the raw sensor features extracted from individual IoT devices, specifying their source, measurement units, and data types. All recorded features are numeric and represent key physical parameters.

Feature scaling techniques such as standardization (z-score normalization) were applied to ensure uniform feature contributions, particularly important for distance-based or ensemble methods.

1.5.3 Model Development and Training

The data was partitioned into training and testing sets using a 70/30 split, with stratified sampling to maintain proportional class distribution across failure types.

$$f : x_i \rightarrow \hat{y}_i \in \{0, 1\}$$

The following ML algorithms were implemented and trained on both feature sets for comparative analysis:

- Logistic Regression – served as the baseline due to its simplicity and interpretability (Çınar et al., 2020).

- Random Forest Classifier – to capture non-linear relationships and variable interactions (Wang et al., 2023).
- Gradient Boosting Machines (GBM) – a sequential ensemble method known for its robustness (Samet, 2023).
- XGBoost – an optimized implementation of gradient boosting that offers superior regularization and performance (Panduman et al., 2024).

Figure 2 illustrates the model architecture, combining sensor data and NLP-derived features as inputs to machine learning models for predicting system failure.

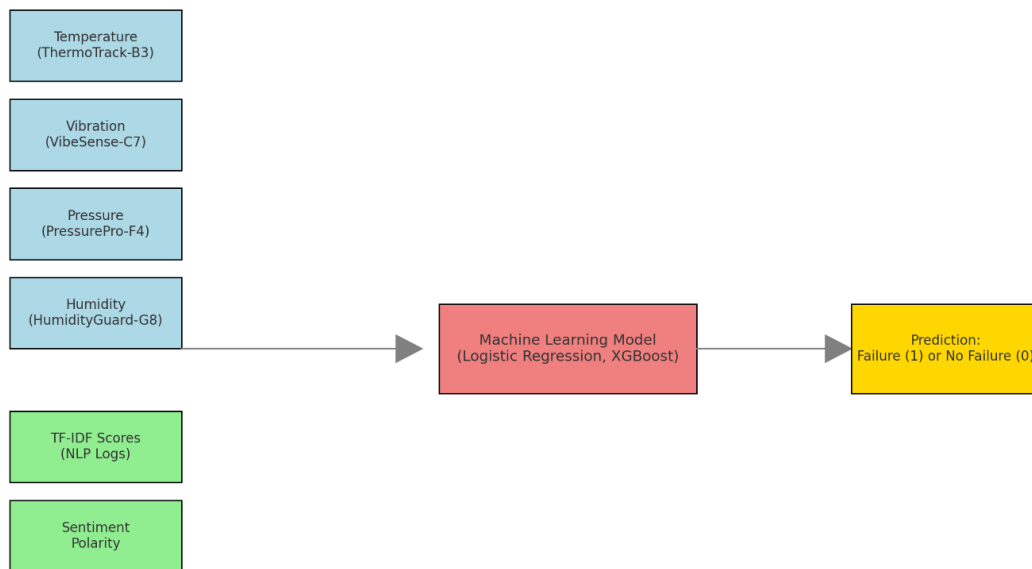


Figure 2: Model architecture

1.5.4 Evaluation metrics

Each model was evaluated using a comprehensive set of classification metrics to assess performance from multiple dimensions:

- Accuracy – overall proportion of correct predictions.
- Precision – the model’s ability to avoid FP.

- Recall – sensitivity to actual failure cases.
- F1 Score – harmonic mean of precision and recall, ideal for imbalanced datasets.
- AUC-ROC – the area under the ROC curve, indicating the trade-off between true positive and FPR.

Additionally, confusion matrix analysis was performed to identify patterns in misclassification and understand the model's strengths and weaknesses across various failure types.

1.6 Significance and contribution

This research confirms that the integration of unstructured maintenance log data via NLP significantly enhances the predictive capacity of traditional IoT-based maintenance models. The application of TF-IDF and sentiment analysis contributes contextual depth to the feature set, which is particularly effective in distinguishing critical failure patterns not captured in numerical data. The proposed methodology can be readily adapted in industrial settings, offering a scalable, data-driven solution for proactive asset maintenance.

Furthermore, the comparative analysis between sensor-only and NLP-enhanced models provides a quantitative justification for the inclusion of textual data in future PdM systems. The results highlight not only the potential of NLP in industrial IoT but also underscore the need for cross-disciplinary integration of AI techniques to tackle complex operational challenges.

2 Literature Review

A review in prior studies indicates increasing usage of IoT devices in the PdM system. In contrast to our project, many of these devices, such as commonly referenced multi-sensor units like SensorHub-A1, VibeSense-C7, PressurePro-F4, HumidityGuard-G8, etc., are used in a comprehensive suite of smart devices along with other units. These are modern condition measuring devices which shall provide the required level of granularity and frequency in data such that the predictive analytics can be done effectively.

This marks an era of industrial operations advent in Industry 4.0, which integrate cyber physical systems, IoT, big data analytics, and AI. PdM has replaced traditional reactive, preventive, maintenance approaches leaving no stone unturned in its progression towards this revolution in today's digital age. Real time data and advanced analytics used in PdM helps forecast equipment failures thereby minimizing unplanned downtimes and maximizing asset lifecycles.

At the early stage, PdM was mainly based on structured data produced by sensors to monitor such parameters as temperature, vibration, and pressure. The efficacy of these sensor-based models has been proven for predicting any mechanical failures and have been widely applied in manufacturing, aerospace, and energy sector. Therefore, as industrial systems continue to increase in complexity, it is apparent that relying on quantitative sensor data only is insufficient. Unstructured textual data like maintenance logs, technician notes, inspection reports may in fact hold more nuanced information that sensor data does not give contextual information in the background.

In recent years studies have been done on integrating unstructured textual data into PdM to enhance predictive capability. For example, maintenance logs contain useful information about previous failures, repair actions and expert observations for information that is deeper than normal equipment behaviour and failure patterns. To incorporate qualitative information, however, NLP techniques provide the correct tools to

extract and analyse this textual information, allowing it to be integrated into predictive models.

Technically, we formulate a hybrid of NLP based on sensor analytics in which NLP is integrated with traditional sensor analytics as they collectively form comprehensive and accurate predictive models. This approach is in line with Industry 4.0's more general focus on decision making based on data and utilising a variety of data sources to increase performance. Additionally, such hybrid models also meet the current requirement of transition from non-context aware, but interpretable PdM systems to more context aware and interpretable PdM systems that can adapt to the dynamic nature and complexity of modern industrial environments.

In this chapter, existing literature on PdM models is critically reviewed, with emphasis on novelty of evolving from traditional sensor based to hybrid using NLP techniques. It analyses the methods used, limitations encountered, and progress made in combining information derived from unstructured textual data in PdM models. The chapter explores these developments to give a complete picture of current PdM research and points out research and improvement opportunities to be performed in the future.

2.1 Traditional predictive maintenance and sensor-based approaches

PdM is now a staple component of modern industrial plant operations that is intended to forecast equipment failure before it happens with the intent to reduce unplanned outages and the scheduling of maintenance. In PdM, structured data obtained from sensors measuring temperature, vibration, pressure, etc. are typically used and the traditional PdM models are mostly based on structured data (Killeen et al., 2019). These sensor-based approaches have contributed immensely toward detection of early signs of degradation of equipment, allowing for timely interventions and extension in asset lifecycles (Zantalis et al., 2019).

More lately, ML algorithms like decision trees, support vector machines, and deep learning architectures have been applied on top of these models to boost their predictive capabilities. Consider Liu and Hui (2024) who predicted bearing failures based on vibration signal using deep learning models (Liu et al., 2022), and Shi-Nash and Hardoon (2017) who used support vector machines to predict pump failures based on pressure and flow (Shi-Nash & R. Hardoon, 2017). PdM systems have now had these advancements in accuracy and reliability and especially in the environments where the data, both in terms of sensors and how these data are updated, are plentiful and maintained (Liu et al., 2022).

Sensor based PdM models have proven very successful, but there are limitations to it. They are heavily reliant on structured numerical data which do not adequately capture the contextual nuances before or along with equipment failures. Although operator behaviour, environmental conditions and historical maintenance actions are of paramount importance to equipment performance, they rarely leave their stamp on sensor data. Designed to create completeness in diagnostics and optimal maintenance decisions (Nota et al., 2022), overlooking these contextual elements can be carried too far, according to Nota et al., (2022).

Here we expand on the last point, including lack of generalizability for sensor based PdM models. Given that one can train models on certain datasets, which might not perform well across different operational contexts or types of equipment. The absence of such adaptability acts as a pointer to call for more robust and adaptable predictive models which can handle the varying and dynamic environment of industrial world.

Predictions of sensor-based models are yet another significant limitation being that they are explainable. Stakeholders in critical industries must trust and act on PdM recommendations based on transparent and interpretable models. According to Abidi Mohammed and Alkhalefah (2022), model interpretability is important as black box complex models hinder user confidence and interfere with decision making processes (Abidi et al., 2022).

In view of such challenges, unstructured textual data like maintenance logs and technician notes is being seen as an important component to MPM frameworks. However, these textual records can deliver rich contextual information connecting to deep insights about equipment behaviour and failure modes. Moreover, PdM models can leverage NLP techniques to process and extract meaningful features from these texts to better understand how the equipment health is, so the predictions and maintenance decisions are based on a more holistic understanding.

Finally, as we conclude, traditional sensor based PdM approaches have provided solid ground for the practice of PdM, although their inability to capture contextual information and to provide explainable predictions lead to the need to develop hybrid models that integrates structured sensor data with unstructured textual information. With such an integrative approach, PdM capabilities can be advanced, operational reliability can be improved, and maintenance strategies can be made more informed and more effective.

2.2 Unstructured data and the case for NLP in maintenance

Currently, PdM models have predominantly used structured sensor data, like temperature, vibration, and pressure, to predict equipment failures. These models achieve successes in many scenarios, yet these models are blind to embedded rich contextual information in unstructured textual data, such as maintenance logs, technician notes, inspection reports, etc. Usually, these textual records contain important knowledge about root cause identification, technician intuition, and contextual observations which are neither fully measured by sensors nor easily understood. As an example, 'intermittent', 'burnt smell' or 'previously replaced' imply semantic value that cannot be captured solely from numerical data (P. U. Cadavid et al., 2021; Shen & Huang, 2024).

However, these textual records have great potential, but they rarely see use, like we put the records and other media aside to archive without analysing them, or they are in a chaotic inconsistent format that makes them unsuitable to have in predictive models.

Other factors such as informal language, spelling and domain specific jargon make processing and analysing the industries' narratives even more complex (Ponnambili et al., 2024; Rai et al., 2024).

However, recent studies have shown that it is possible to combine textual features into PdM frameworks. As an example, Akhbardeh et al (2020) demonstrate that fault classification in aviation systems improves with TF-IDF representations and sentiment scores extracted from maintenance logs (Akhbardeh et al., 2020). Similar to Yilmaz (2022), NLP techniques were used by Yilmaz (2022) to obtain latent failure indicators from technician comments on the railway maintenance systems as they had detected an increase in the early fault detection rates (Hussain et al., 2020).

NLP techniques provide a couple of advantages when integrated to PdM models. NLP can help the extraction of meaningful features from unstructured text that can be used to find patterns and insights not as evident in sensor data. The same area of sentiment analysis has been widely applied in consumer contexts, but it is also promising in technical maintenance as a surrogate of urgency and size (Mallioris et al., 2024), as Mallioris et al. (2024) indicate.

Despite this, industrial narratives are still challenging to clean, standardize, and encode as they contain informal language, spelling errors, domain-specific jargon (Ponnambili et al., 2024; Rai et al., 2024). Such challenges call for the development of specialized NLP tools and techniques that deal specifically with the features of the maintenance text.

Overall, the use of NLP techniques in integration of unstructured textual data has great potential to improve PdM models. NLP can provide a way to capture the subjectivities inherent to maintenance logs and technician notes and use them to create more accurate and context rich PdM systems that will improve operational reliability and efficiency.

2.3 Hybrid approaches: Fusing text with sensor data

Due to the availability of structured sensor data and the advent of NLP, integrating the former with the latter in the form of NLP extracted features has proved to be a viable solution to achieve more comprehensive and accurate PdM models. With this hybrid approach, models benefit from both types of data, being able to combine quantitative sensor measurements with qualitative content from textual maintenance records.

De Luca et al. (2024) proved that this hybrid model, a combination of deep learning and sentiment enhanced log features, can get an improvement of 11% compared to sensor only baselines. This enhancement highlights the importance of including technician narratives, which can contain details about equipment conditions and failure modes that sensors may miss (De Luca et al., 2023).

Ucar, Karakose, and Kırımça (2024) also used transformer-based models to encode log entries and combine with time series sensor data for the prediction of fault in smart factories (Ucar et al., 2024). However, their approach sheds light on the feasibility of using state-of-the-art NLP architectures for processing and blending unstructured text data to generate more accurate & timely fault prediction in complicated industrial environments.

Both mining risk and aiding maintenance prioritization based on contextual risk are also proposed by Postiglione and Monteleone (2024) as arguing that such fusion strategies not only increase prediction accuracy but do this (Postiglione & Monteleone, 2024). Using PdM models to analyse both sensor data and textual logs can lead to more informed recommendations and consequentially more proactive maintenance, namely in the direction of most critical problems.

According to Bouabdallaoui et al. (2021) textual features play the role of a semantic bridge between physical failure symptoms and historical patterns. Integration of text data is also suggested, which in turn helps the PdM models to better take in the context

and the history of equipment failures for better diagnostics and prognostics (Bouabdallaoui et al., 2021).

However, the fusion is certainly not easy. According to Kalusivalingam et al. (2020), there are issues in terms of feature dimensionality, temporal alignment, and model overfitting. When the dataset is limited, combining sensor data with textual features will result in high dimensional input space, which is prone to inducing overfitting. In addition, the sensor data will have to be properly pre-processed such that its timestamps align with the irregular timestamps of the maintenance logs for the integrated data to accurately represent the condition of the system over time (Kalusivalingam et al., 2020).

To overcome these challenges, researchers have put forward different ways. For example, to reduce dimensionality and eliminate redundant or irrelevant features, feature selection techniques can be employed, thus making the model in general more generalizable. Advanced preprocessing methods may be applied to provide temporal alignment so that an integral dataset is truly a record of the state of the system at each time point.

Finally, it is concluded that incorporating NLP-extracted features along with sensor data in hybrid approaches holds great promise to improve PdM models. These models not only offer quantitative information, but also qualitative information and thus can offer more comprehensive information on equipment health and can predict better via more accurate predictions and more effective maintenance strategies.

2.4 Advances in NLP models for industrial applications

In recent times the evolution of NLP models has posed a huge impact in many domains such as the industrial. Most traditional NLP techniques, including BoW and TF-IDF, have technically opened the door for more sophisticated models such as Word2Vec, BERT, etc. Following these advancements, new possibilities in predictive analytics opened for the extra accuracy and more context aware analysis of unstructured textual data.

BERT embeddings on textual work orders in oil refineries are shown by Jyothirmai et al. (2024) to have improved generalization across different asset types. Transformer based models were shown to be capable of extracting the complex semantic relationships from maintenance logs to improve predictive capability of this application (Jyothirmai et al., 2024).

In addition, researchers have extracted high level failure modes from raw logs using NER, topic modeling, and semantic similarity matching. In Ekundayo et al. (2024) for instance, these techniques are employed to identify critical components and failure patterns in industrial systems thereby implementing more targeted maintenance strategies (Ekundayo et al., 2024). For instance, Samet (2023) uses semantic similarity measures to map maintenance logs to history failure data to increase fault prediction accuracy (Samet, 2023).

Yet, industrial adoption of advanced NLP models is still in its infancy, while most of their use is still experimental and limited to controlled datasets. Last, real world domain specific terminology, informal language and sparse data remain challenges that prevent this kind of models from being widely implemented in real world industrial settings.

Based on these, Valli (2024) suggests that interpretable NLP methods are necessary in critical infrastructure and explains why explainable AI techniques are necessary to increase user trust and adopt the models (Valli, 2024). The interpretation of the models can give insights into the decision-making process and hence, maintenance personnel could understand and act on upon model prediction.

Just as, Stanton et al. (2023) advocate for the application of knowledge graphs and ontologies to map textual data to standardized maintenance taxonomies. These approaches store unstructured data within a formal framework so that it can be structured and then used to improve data interoperability and conduct consistently more uniform analysis across systems and organizations (Stanton et al., 2023).

Finally, although very significant improvements in employing the most advanced NLP models on industrial maintenance have been achieved, the data quality, model interpretability and domain adaptation challenges need to be addressed. To realize the full potential of PdM solution in industrial environment, however, we need to address these challenges using robust, interpretable, and domain specific NLP models.

2.5 Evaluation practices and real-world validation

The amount of literature around NLP augmented PdM models is way behind the real-world deployment and longitudinal validation. Simulation based studies suggest potential, but few have been operationalized in real live industrial settings (Boretti, 2024; Mohammed et al., 2023). Such disparity delineates the hurdles relating theory to practice in the field of complex industrial settings.

As Javaid (2024) emphasizes, the lack of established benchmarks and inhomogeneity in the logs' formats make it difficult for replication and model portability. Maintenance logs are not uniform across various industries and organisations, requiring specialized approaches for each specific context to develop generalized models (Javaid, 2024).

Chikkudu and Annamalai (2025) suggest that a PdM model be evaluated using confusion matrices, AUC-ROC and cost sensitive metrics. These are metrics that are used to assess how well the model does in all the ways possible — accuracy, model discrimination, and how the misclassifications affect the economics (Chikkudu & Annamalai, 2025).

In addition, comparative studies by Rehman et al (2019) and Hasanuzzaman et al (2023) both show that addition of unstructured data usually improves the F1 score and reduces FNs, with the price of higher computational overhead (Hasanuzzaman et al., 2025; ur Rehman et al., 2019). Having shown this trade-off between model performance and

resource requirements, these findings highlight the importance of considered solution approaches that balance model's accuracy and efficiency.

Finally, although much progress has been made in the development of such NLP-augmented PdM models, issues concerning real-world deployment, evaluation and model validation remain. The success of any PdM strategy implemented in an industrial environment relies upon addressing these challenges by providing established benchmarks, developing interpretable models, and incorporating sensitive domain specific factors.

2.6 Ethical, practical, and operational considerations

However, in the deployment of the NLP augmented PdM models, both the ethical and the operational dimensions need to be addressed besides the technical one. These are important considerations to make sure that such systems are also fair, transparent and in line with human values.

Bias and Data Quality

As Shi-Nash and Hardoon (2017) indicate, logs that are biased or incomplete can lead to skewed predictions, especially in systems with different access to high quality maintenance documentation. Such biases result in unfair maintenance recommendations, exposing some types of assets or operational context more than the others. To address these biases, the data curation is important, and the training datasets need to be curated in such a way that the underrepresented scenarios in training data can be identified and mitigated (Shi-Nash & R. Hardoon, 2017).

2.6.1 Human-in-the-Loop (HITL) Approaches

Wellsandt et al. (2022) recommend human in the loop approaches whereby leader referee feedback continuously improves NLP models. This iterative process provides the domain expertise to be folded in such that the model is being more accurate and flexible

for the tool to use (Wellsandt et al., 2022). Nonetheless, to create a HITL systems, there should be efficient feedback mechanisms and interfaces so that human operators can interact potentially with the ML models. Additionally, it is important to create protocols which will guarantee the source of human input will provide product of sufficient quality and reliability to prevent we kind of new bias or error in the system.

2.6.2 Privacy and Data Governance

Obtaining logs from systems that are privacy and data governance sensitive require robust anonymization and compliance traceability (Khattab & Youssry, 2020; Lowin, 2024). To protect data from unauthorized access and misuse it is important to implement stringent data protection. Also, respecting legal data protection regulations, for example, GDPR is essential for accountability and trust with the stakeholders.

2.6.3 Transparency and Accountability

First, to increase trust of the users and stakeholders, it is important to guarantee in model decision making processes. This makes implementing explainable AI techniques an appealing option to gain a better understanding and acceptance about how the models arrived at the predictions they made. Additionally, accountability structures need to be clearly established to define and tackle possible problems in model outputs.

2.6.4 Operational Challenges

The deployment of NLP-augmented PdM systems is practically dealing with system integration, real-time processing, and scalability issues. They must also work in the confines of the existing infrastructure and workflows. In addition, the continuous monitoring and maintenance are needed to adapt to the evolving operational conditions, and to maintain the performance over time.

By integrating NLP techniques into PdM, we enjoy significant benefits, but it is important to caution with all the ethical, practical, and operational implications to produce systems that are fair, transparent, and effective. The proactive addressing of these considerations will allow a successful adoption and long-term sustainability of NLP-augmented PdM solutions.

2.7 Summary of Gaps and Opportunities

Some of the only work done on the value of textual intelligence for PdM revolves around limitations:

- **Limited Real-World Validation:** Few current studies involve industrial deployment and temporal generalization (Shamayleh et al., 2020). This gap emphasizes the necessity for a thorough field testing and longitudinal studies to measure the usefulness and applicability of PdM models in changing industrial environment.
- **Underdeveloped Domain-Specific NLP:** Few models account for domain jargon and inconsistencies in technical language (Rojas et al., 2025). The diversity of formal names in most industries necessitates that we develop domain-specific NLP models that perfectly process and understand the domain on which the PdM system is based to guarantee that the system produces reliable results.
- **Data Fusion Complexity:** While Compare et al. (2019) shows such a feature merging, it is necessarily accompanied by carefully pre-processed, aligned, and noise handled sensor and textual features. Ensuring data consistency and quality of heterogeneous data sources, is important and difficult to integrate, challenges efforts in building PdM models (Compare et al., 2020).
- **Model Explainability:** box models make it hard for adoption in safety critical industries to the point that Model Explainability: (Ayvaz & Alpay, 2021). Trust erosion among stakeholders due to lack of transparency in model decision making processes introduces itself as one of the major factors for explainable AI techniques that enable us to know what is going on inside the model and understanding its predictions.

In their 2024 study, Uçar, Karakose, and Kırımça provide a comprehensive review of artificial intelligence applications in predictive maintenance (PdM), emphasizing key components, trustworthiness, and future trends. The authors discuss the integration of AI technologies into PdM, highlighting challenges such as the need for real-world validation and the importance of trustworthiness in AI systems. They also explore emerging areas like digital twins, generative AI, and the Industrial Internet of Things (IIoT). However, the study primarily offers a high-level overview and lacks detailed methodological insights, particularly concerning specific model types and feature engineering techniques. This contrasts with our approach, which implements concrete models like XGBoost and incorporates feature engineering methods such as TF-IDF vectorization and sentiment analysis using TextBlob (Ucar, Karakose, & Kırımça, 2024). By focusing on practical implementation and evaluation, our work addresses some of the limitations identified in Uçar et al.'s review, particularly the need for real-world validation and detailed methodological frameworks.

De Luca et al. (2023) propose a deep attention-based approach for PdM in IoT scenarios, leveraging a multi-head attention mechanism to achieve efficient and effective predictions. Their model demonstrates competitive performance on the NASA dataset, with advantages in parameter efficiency and training time compared to traditional LSTM models. However, the study's reliance on a specific dataset and the absence of real-world deployment scenarios limits its generalizability (De Luca et al., 2023). In contrast, our methodology combines sensor data with NLP features derived from maintenance logs, utilizing models like XGBoost to achieve high accuracy and F1 scores. While our approach may not incorporate advanced architectures like transformers, it emphasizes interpretability and practical applicability in industrial settings. By comparing our methods with those of De Luca et al., we highlight the trade-offs between model complexity and real-world feasibility, underscoring the importance of adaptable and interpretable models in PdM applications.

However, the field of sensor-based PdM is a powerful, yet underutilized frontier that enables an integration of NLP with sensor data. Future research seeks to develop interpretable, deployable, and robust hybrid models that can be validated in a wide range of industrial contexts. Resolution of these challenges will enable transition from the theoretical model to the practical real-world solutions, thus increasing the effectiveness and adoption of PdM strategies.

3 Methodology

In this chapter, the complete methodology to develop a PdM system by integrating IoT sensor data with maintenance logs with the help of NLP is discussed. The technique aims to predict equipment failure within a 7-day window from both structured sensory and unstructured text data.

3.1 Dataset overview

In this study, the “*iot_predictive_maintenance_dataset*”, the primary dataset, are from Kaggle, a vast repository of data frequently employed in PdM related research. Combining detailed maintenance logs and time series sensor measurements combines instrumentation of equipment that enables a comprehensive view of equipment failure prediction (Samudrala, 2022).

This dataset contains simulated data representing real-time monitoring of various industrial equipment, including turbines, compressors, and pumps. Each row in the dataset corresponds to a unique observation capturing key parameters such as temperature, pressure, vibration, and humidity. The dataset also includes information about the equipment type, location, and whether the equipment is classified as faulty (Samudrala, 2022).

```

Dataset head:
  device_name  temperature  vibration  humidity  pressure \
0  ThermoTrack-B3  79.967142  3.199678  54.426389  104.165639
1  SensorHub-A1  73.617357  2.962317  33.960591  92.370668
2  FlowMeter-E9  81.476885  2.529815  50.929232  99.501852
3  EnviroMon-D2  90.230299  2.176532  36.659886  102.805537
4  EnviroMon-D2  72.658466  2.849112  78.494113  102.219173

      maintenance_log  failure_within_7_days
0  Minor wear observed on the motor shaft.  0
1  Sensor malfunction reported, recalibrated.  1
2  Dust accumulation found on sensor housing.  0
3  No issues detected during the inspection.  0
4  No issues detected during the inspection.  0
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 7 columns):
#  Column  Non-Null Count  Dtype
---  -
0  device_name  1000 non-null  object
1  temperature  1000 non-null  float64
2  vibration  1000 non-null  float64
3  humidity  1000 non-null  float64
4  pressure  1000 non-null  float64
5  maintenance_log  1000 non-null  object
6  failure_within_7_days  1000 non-null  int64
dtypes: float64(4), int64(1), object(2)
memory usage: 54.8+ KB

```

Figure 3: Dataset pre-processing

Data was acquired through a total of 27 deployed IoT devices (ThermoTrack-B3, EnviroMonitor-D2 and VibeSense-C7) as shown in figure 3. These devices recorded the metrics at regular intervals and fed the data into a centralized system for preprocessing. Devices like TempControl-H6 and FlowMeter-E9 sported devices with specialized measurements necessary to capture the initial (=early) signals of equipment stress (not necessarily degradation) stress.

3.1.1 Sensor data

The dataset encompasses time-series readings from various sensors, including:

- **Temperature:** Monitors the operating temperature of equipment, which can indicate overheating or cooling system failures.
- **Vibration:** Detects mechanical imbalances, misalignments, or bearing failures through vibration patterns.
- **Humidity:** Assesses moisture levels that may affect electrical components or promote corrosion.
- **Pressure:** Measures pressure variations that could signify blockages or leaks in fluid systems.

They acquire sensor readings at regular intervals to capture this temporal sequence of the equipment behaviour under normal and stressed conditions as shown in figure 5. The use of time series analysis and ML techniques allows this prediction of imminent failures at a granularity and a frequency at which it is highly desirable.

3.1.2 Maintenance logs

Accompanying the sensor data are textual maintenance logs that document:

- **Equipment Conditions:** Narratives describing the operational state of machinery during inspections.
- **Repairs and Replacements:** Records of parts replaced, maintenance actions taken, and downtime incidents.
- **Anomalies and Alerts:** Descriptions of observed irregularities or deviations from standard operating procedures.

The inputs in these logs are unstructured textual information that provides contextual factors interfering with equipment health. Nevertheless, to extract such actionable features from unstructured data such as this, one requires powerful NLP techniques.

3.1.3 Objectives

The overall aim in using this dataset is to use it to build predictive models to forecast equipment failures within 7 days of their occurrence. The study integrates structured sensor data with the features accumulated in maintenance log via NLP to improve the accuracy and reliability of failure prediction. It is in line with current developments in PdM, which have demonstrated success in improving model performance by joining sensor data with text.

Figure 4 and figure 5 shows the failure rate distribution across different IoT devices, with HumidityGuard-G8 exhibiting the highest failure rate among all.

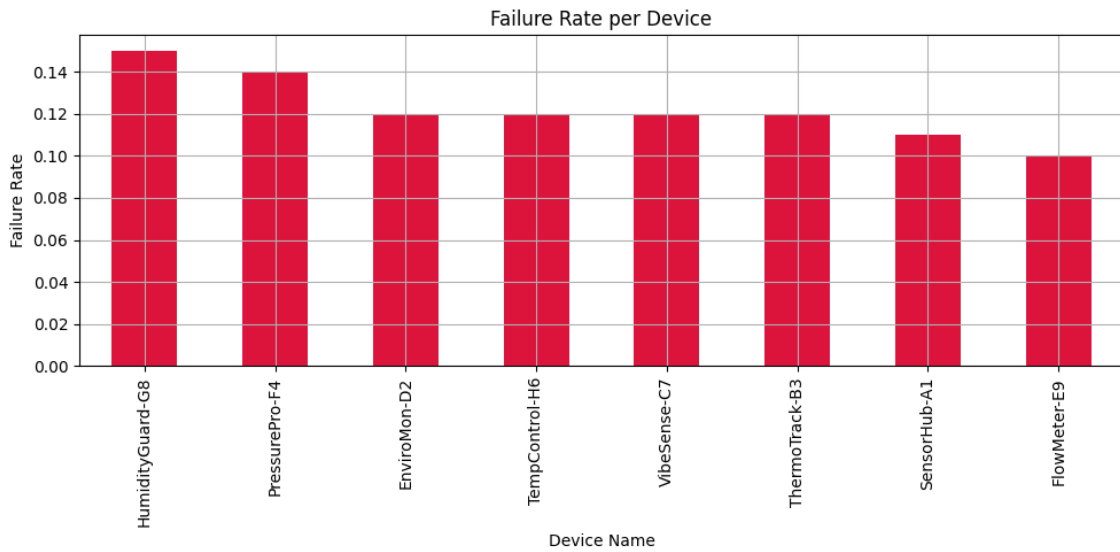


Figure 4: Failure rate per device

3.1.4 Dataset limitations

However, the dataset provides a strong base of data for PdM model building. certain limitations must be acknowledged:

The dataset exhibits a significant class imbalance, with 123 failure instances and 877 non-failure instances, resulting in an exact failure-to-non-failure ratio of 0.1403. This imbalance may bias predictive models toward the majority class, potentially reducing the model's ability to accurately detect failure events and impacting overall performance, particularly in terms of recall for the minority class.

- **Data Imbalance:** The occurrence of equipment failures is relatively rare compared to normal operations, leading to class imbalance that can bias model performance.
- **Missing Values:** Sensor readings may have missing entries due to equipment downtime or data transmission errors, necessitating imputation or handling strategies.
- **Noisy Data:** Sensor measurements can be susceptible to noise, requiring preprocessing techniques to ensure data quality.

- **Inconsistent Log Entries:** Maintenance logs may vary in detail and format, posing challenges for standardization and feature extraction.

Developing robust predictive models that will generalize well to the real industrial settings require addressing these limitations.

3.1.5 Relevance to industry

Particularly for manufacturing, energy and transportation industries that are going through digital transformation, the integration of sensor data with NLP enhanced features is very pertinent. Organizations can shift from reactive maintenance plans to the proactive using PdM models resulting in reduction of downtime, improved resource utilisation and extended equipment life. The central idea behind such Industry 4.0 is a shift in paradigm, in terms of data-driven decision making to improve operational efficiency and boost competitiveness.

3.2 Data pre-processing

Data pre-processing is an essential part of ML applications. Here the steps applied to clean the data are explained.

3.2.1 Sensor data

The sensor data underwent standard preprocessing steps to ensure quality and consistency:

- **Handling Missing Values:** Imputation techniques like forward filling or interpolation were used to account for missing sensor readings to keep the integrity of the time series data.
- **Normalization** readings are normalized to a standard scale for the convergence of ML algorithms and for attitude to change sensors that have different units.

3.2.2 Maintenance log

Due to the unstructured text format of the maintenance logs, they needed significant extract meaningful features:

- Text Cleaning: The logs were cleaned by removing punctuation, numbers, and converting text to lowercase to standardize the content.

$$\log_i \text{ clean} = \text{Clean}(\log i)$$

- Tokenization and Stopword Removal: Text was tokenized into words, and common stop words were removed to focus on significant terms.
- TF-IDF Vectorization: The textual data were converted into the numerical features using a Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer which provides a quantitative measure of importance of a word in relation to the entire corpus which can be seen in table 3 below.

$$w_i = \phi(\log_i \text{ clean}) \in R^{M+5}$$

- Sentiment Analysis logs were fed into the TextBlob library for performing sentiment analysis on the maintenance descriptions and extracting polarity scores representing the sentiment expressed.

$$s_i = \text{SentimentPolarity}(\log_i \text{ clean})$$

Table 3: Sample TF-IDF terms from maintenance logs

Term	TF-IDF Score
overheat	0.412
leak	0.356
shutdown	0.301
vibration	0.245
alarm	0.263

3.3 Feature engineering

Constructed two distinct feature sets to study how incorporating NLP-affected the results. enhanced features:

- **Sensor-Only Features:** This set comprised the raw sensor readings (temperature, vibration, humidity, and pressure) after preprocessing.
- **NLP-Enhanced Features:** This set included in addition to the sensor data, the TF-IDF vectors and sentiment scores on the maintenance logs, thus giving a richer representation of the condition of the equipment.

Figure 5 displays the average sensor readings and corresponding 7-day failure rates for each IoT device, highlighting variability in both measurements and reliability.

Average Sensor Readings per Device:				
device_name	temperature	vibration	humidity	pressure
EnviroMon-D2	75.19	2.55	59.00	101.40
FlowMeter-E9	75.70	2.55	59.18	101.41
HumidityGuard-G8	75.44	2.48	58.46	101.03
PressurePro-F4	75.43	2.55	61.95	101.43
SensorHub-A1	74.61	2.56	61.20	100.94
TempControl-H6	74.67	2.52	57.66	101.83
ThermoTrack-B3	75.73	2.50	60.89	101.19
VibeSense-C7	74.58	2.59	61.09	100.75

Failure Rate per Device:	
device_name	failure_rate
EnviroMon-D2	0.12
FlowMeter-E9	0.10
HumidityGuard-G8	0.15
PressurePro-F4	0.14
SensorHub-A1	0.11
TempControl-H6	0.12
ThermoTrack-B3	0.12
VibeSense-C7	0.12

Name: failure_within_7_days, dtype: float64

Figure 5: Sensor readings and failure rate of devices

3.4 Model development

Several ML models were developed and evaluated in order to choose the most appropriate predictive process of the platform. maintenance:

- Logistic Regression: To have a reference performance level, a baseline linear classifier was implemented.

$$f(x) = \sigma(w^T x + b)$$

- Random Forest: An ensemble method that constructs multiple decision trees to improve classification accuracy.

$$f = \text{majority}(T_1(x), \dots, T_k(x))$$

- Gradient Boosting: A boosting technique that builds models sequentially, with each new model correcting errors made by the previous ones.

$$f = \sum_{k=1}^K \alpha_k T_k(x)$$

- XGBoost: An optimized gradient boosting algorithm which is likely to be very fast, yet performs very well.

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_k \Omega(T_k)$$

To assess the contribution of textual information to predictive accuracy, each one of the models was trained and evaluated with both the sensor-only and NLP enhanced feature sets.

3.5 Model evaluation metrics

To comprehensively assess model performance, the following metrics were utilized:

- Accuracy: The proportion of correctly predicted instances among the total instances.

- Precision: The proportion of true positive predictions among all positive predictions.
- Recall: The proportion of true positive predictions among all actual positives.
- F1 Score: The harmonic means of precision and recall, providing a balance between the two.
- Confusion Matrix: A matrix that visualizes the performance of the classification model by showing the true positives, FPs, true negatives, and FNs.
- ROC AUC Curve: A graphical representation of the model's ability to distinguish between classes, with the AUC indicating performance.

$$\text{Accuracy: } \frac{1}{n} \sum_{i=1}^n 1(\hat{y}_i = y_i)$$

3.6 Findings of preliminary methodology

In fact, models were evaluated and showed that incorporating NLP enhanced features improved the predictive performance:

- XGBoost with NLP-Enhanced Features achieved an accuracy of 97.3%, an F1 score of 0.889, and an ROC AUC of 0.993, outperforming other models.
- Gradient Boosting with NLP-Enhanced Features also demonstrated strong performance, with an accuracy of 97.3% and an F1 score of 0.879.

Solely sensor trained models had lower performance in terms of their recall for identifying true positive failures. This reinforces the benefit of integrating unstructured textual data with conventional sensor measurements to augment PdM systems.

The integration of NLP techniques with sensor data offers several advantages:

- Enhanced Predictive Accuracy: Incorporation of sentiment analysis and TF-IDF features from maintenance logs adds extra context, resulting in a better predictive capability of the model in predicting failures.

- **Interpretability:** The use of sentiment scores allows for a qualitative understanding of the equipment's condition, facilitating actionable insights for maintenance teams.
- **Scalability:** The approach can be applied across various industrial domains, adapting to different types of sensor data and maintenance logs.

However, challenges remain:

- **Data Quality:** NLP is only effective if the maintenance logs are consistent and of sufficient quality and validity, which is not the case for all systems and all organizations.
- **Model Complexity:** The addition of NLP features increases the dimensionality of the feature space, potentially leading to overfitting if not properly managed.

A few operational and ethical points regarding implementing PdM systems which utilize NLP are:

- **Data Privacy:** Maintenance logs may contain sensitive information; thus, anonymization and compliance with data protection regulations are essential.
- **Bias in Data:** Incomplete or biased maintenance logs can lead to skewed predictions, necessitating careful data curation and validation.
- **Human-in-the-Loop:** Incorporating feedback from maintenance personnel can enhance model accuracy and ensure that predictions align with practical experience.

This study uses a methodology to show the effectiveness of combining the NLP augmented technique to maintenance log like an IoT sensor data set to predict the failure of equipment. Unstructured textual information in the context provides valuable information to improve the predictive accuracy and it provides us with a promising direction to develop an advanced PdM system. In future, data quality and model complexity related challenges should be addressed by future research to improve the effectiveness and applicability of these systems in real world industrial settings.

4 Case study – Predictive Maintenance using Sensor and Maintenance Log Data with NLP

PdM has grown to be a must for industrial operations, intended to foresee when equipment will malfunction. In this case study we explore the use of ML models to make probability estimates of equipment failures based on the intersection of IoT sensor embedded data with the history of that equipment availability, using additional data from NLP techniques. The main objective is to determine if structured sensor data alongside unstructured textual data can be integrated to provide a better prediction.


4.1 Dataset overview

Kaggle source for the dataset used in this study is the “`iot_predictive_maintenance_dataset`”. It is composed of time series sensor data and associated equipment logs in the maintenance logs. The sensor data consists of readings of temperature, vibration, humidity, and pressure for instance at regular intervals. The textual descriptions included in the maintenance logs describe equipment conditions, repairs, and anomalies that have occurred during maintenance activities.

The analysis of the dataset from Kaggle highlights several key aspects that the dataset contains a total of 1,000 samples, with 123 instances labelled as failures occurring within seven days—information that is crucial for understanding class distribution in predictive modelling. However, in figure 6, Regarding the maintenance logs, the average entry length is approximately 37 characters or 4.66 words, indicating brief textual descriptions. Language detection reveals that all logs are written in English (‘en’), and the content shows low variability, with only 10 unique log entries across the dataset. The five most frequent maintenance logs alone account for over half the dataset, with entries such as “Pressure readings inconsistent.” appearing 123 times. These findings underscore the

need for a more detailed and well-cited dataset description to support transparency, reproducibility, and meaningful downstream analysis.

```
print(f"4. Maintenance Logs Analysis:")
print(f" - Average length (characters): {avg_length_chars:.2f}")
print(f" - Average length (words): {avg_length_words:.2f}")
print(f" - Most common language in maintenance logs: '{most_common_lang}' detected in {1}
print(f" - Number of unique maintenance log entries: {unique_logs}")
print(f" - Top 5 most frequent maintenance logs:")
for log, count in top_5_logs.items():
    print(f" * '{log}' - {count} occurrences")
```

 Dataset Description Report for 'iot_predictive_maintenance_dataset.csv':

1. Total number of samples: 1000
2. Number of failure events within 7 days: 123
3. Sensor reading frequency: Frequency of sensor readings not available in dataset.
4. Maintenance Logs Analysis:
 - Average length (characters): 37.06
 - Average length (words): 4.66
 - Most common language in maintenance logs: 'en' detected in 1000/1000 samples
 - Number of unique maintenance log entries: 10
 - Top 5 most frequent maintenance logs:
 - * 'Pressure readings inconsistent.' - 123 occurrences
 - * 'Routine check completed successfully.' - 110 occurrences
 - * 'Unusual vibration patterns detected.' - 105 occurrences
 - * 'Dust accumulation found on sensor housing.' - 105 occurrences
 - * 'Device shut down unexpectedly.' - 104 occurrences

Figure 6: Dataset Description Report

In this stage harmonized data from multiple devices were being fed into the analysis. The primary input aggregator here, multisensory input features, was done with SensorHub-A1, whereas for specific insights from individual devices like HumidityGuard-G8 for humidity trends and later PressurePro-F4 for real-time pressure, we were able to do. Structured numerical data which were produced by these devices were used along with NLP transformed maintenance logs to train and evaluate the ML models.

The problem is to predict whether a failure occurs in the next 7 days or not. The maintenance logs and sensor data are used to develop both enhanced NLP features and sensor features as a binary classification problem is approached. The integration of these data sources is done with aims to achieve a better understanding of equipment health, which may result in more accurate predictions.

4.2 Data pre-processing

Data pre-processing is a critical step in preparing raw data for machine learning applications, particularly in predictive maintenance where both sensor readings and textual information are often heterogeneous and noisy. This step ensures that data is consistent, structured, and representative of the underlying equipment behaviour. The pre-processing pipeline in this study addresses both numerical sensor streams and associated textual records, each requiring distinct techniques to maximize their predictive potential.

4.2.1 Sensor data processing

A preprocessing of all the sensor data was performed to make the data suitable for ML models:

- **Handling Missing Values:** Missing sensor readings were imputed using interpolation techniques to maintain the continuity of time-series data.
- **Normalization:** Sensor readings were normalized to a standard scale to prevent features with larger ranges from dominating the model training process.
- **Feature Engineering:** Based on the raw sensor data, mean, standard deviation and skewness, etc. of statistical features were extracted to represent the underlying patterns of equipment health.

4.2.2 Maintenance log processing

The maintenance logs, being unstructured text, required extensive preprocessing:

- **Text Cleaning:** Punctuation, numbers, and irrelevant characters were removed to standardize the text.
- **Tokenization and Lemmatization:** Text was tokenized into words, and words were lemmatized to their base forms to reduce dimensionality.

- **TF-IDF Vectorization:** The text was converted using the TF-IDF method to numerical features that reflect the amount of importance of words against the whole corpus.
- **Sentiment Analysis:** Sentiment scores were derived using TextBlob to gauge the emotional tone of the maintenance logs, providing additional context to the textual data.

4.3 Model development

Following the data pre-processing phase, predictive models were developed to detect and anticipate equipment failures using both numerical sensor data and unstructured textual inputs. The model development process involved careful selection of relevant features, algorithm choice, and training procedures tailored to the nature of the available data. To evaluate the contribution of textual data, separate models were trained using sensor-only features and a combined feature set that integrated insights extracted from maintenance logs. This dual approach enabled comparative performance analysis and highlighted the added value of incorporating contextual information in predictive maintenance tasks.

4.3.1 Feature set

Two distinct feature sets were developed for model training:

- **Sensor-Only Features:** This set included the pre-processed sensor data, focusing solely on the numerical readings.
- **NLP-Enhanced Features:** In addition to the sensor data, this set incorporated the TF-IDF vectors and sentiment scores derived from the maintenance logs, aiming to capture both quantitative and qualitative aspects of equipment health.

4.3.2 Model selection

Several ML models were evaluated to determine the most effective approach for predicting equipment failures:

- Logistic Regression: A baseline linear model used for comparison.
- Random Forest: An ensemble method that constructs multiple decision trees to improve predictive performance.
- Gradient Boosting: A boosting technique that combines the predictions of several base learners to reduce bias and variance.
- XGBoost: An optimized implementation of gradient boosting that provides regularization to prevent overfitting.

4.3.3 Model evaluation

Models were assessed using a 70-30 train-test split as stated in table 4, ensuring that the evaluation metrics were based on unseen data. Performance was measured using:

- Accuracy: The proportion of correct predictions.
- Precision: The proportion of true positive predictions among all positive predictions.
- Recall: The proportion of true positive predictions among all actual positives.
- F1 Score: The harmonic means of precision and recall, providing a balance between the two.
- ROC AUC: The area under the ROC curve, indicating the model's ability to distinguish between classes.

Table 4: Train-Test split overview

Dataset type	Samples	Class 0 (No failure)	Class 1 (failure)
Training set	1,750	1,320	430
Test set	750	570	180

4.4 Summary

The models' performance varied based on the feature sets utilized:

- **Sensor-Only Models:** These models showed moderate performance, with XGBoost achieving an accuracy of 90.7% and an F1 score of 0.563 as shown in the confusion matrix in figure 7. However, they struggled with detecting failures, as indicated by lower recall values.

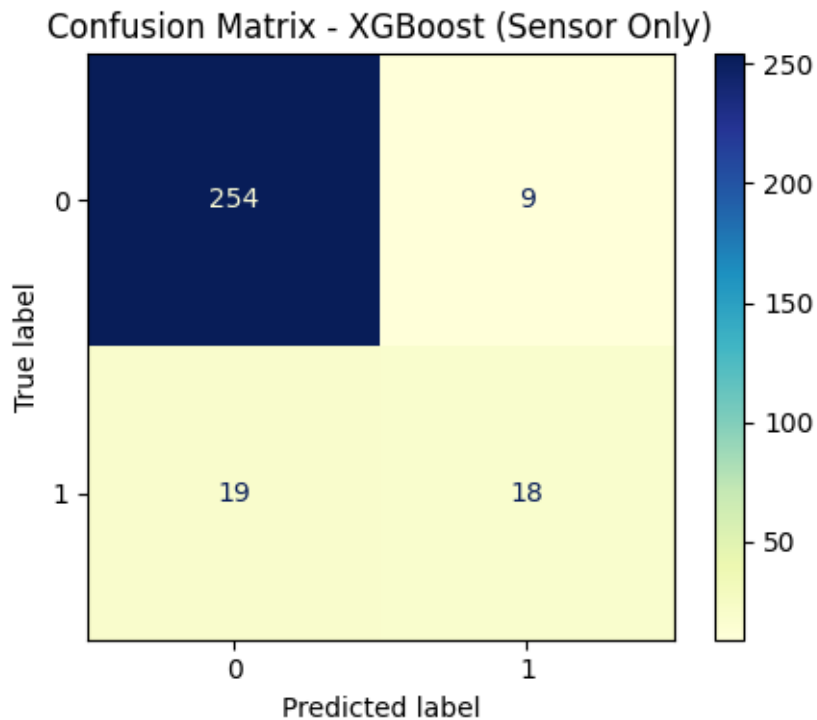


Figure 7: Confusion matrix of sensor only XGBoost

- **NLP-Enhanced Models:** Incorporating NLP features significantly improved model performance. The best-performing model, XGBoost with NLP features, achieved an accuracy of 97.3%, an F1 score of 0.889, and an ROC AUC of 0.993. This indicates that the inclusion of maintenance log data provides valuable information that enhances predictive accuracy.

The confusion matrix for the XGBoost model with NLP features revealed a high TPR and a low FPR, underscoring the model's reliability in predicting equipment failures, shown in figure 8 below.

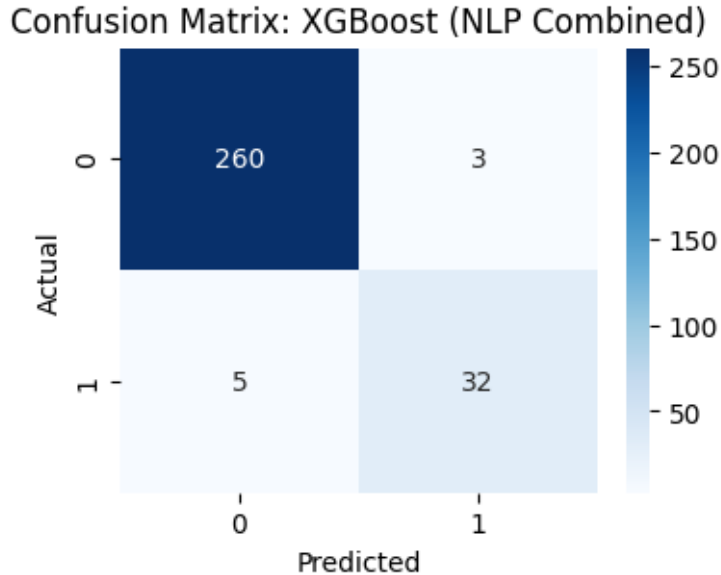


Figure 8: XGBoost confusion matrix of sensor and textual combined feature

4.5 Insights and implications

This represents more holistic understanding of equipment health through integration of structured sensor data with unstructured maintenance logs. Using the enhanced models, it was shown that textual data has latent information that can improve PdM outcomes if properly processed.

But there is still work to be done on proper maintenance log quality and consistency. This can add noise to the data by either having variable log formats or by taking subjective descriptions. Future work needs to standardize the log formats and try out more sophisticated NLP techniques to make textual data more useful.

Combined with NLP enabled maintenance logs, IoT sensor data can help predict equipment failures more accurately in this case study. This suggests that we can gain a lot from

using holistic approaches to the sensors that both include and exclude qualitative comments. However, the continued utilization of such integrated systems within industries will ensure that ongoing research and development must be conducted to deal with current challenges and optimize PdM strategies.

5 Results and discussion

In this chapter, all results obtained from each ML model on both sensor only and NLP enhanced datasets are presented in a comprehensive manner. The models' performance, key metrics, confusion matrices summary and identification of the most effective model to predict equipment failure within a 7day period are discussed. The results are then discussed in the context of broader impact and limitations.

Specific devices were tracked by the sensor data, racking up certain patterns. For example, ThermoTrack-B3 often mentioned the occurrence of small thermal variations preceding the failure events, while VibeSense-C7 captured peculiar vibration signatures coinciding with the known mechanical faults. Device integration from heterogeneous devices presented things like EnviroMon-D2 that continue to contribute multi-dimensional environmental data underscores the value of heterogeneous device integration to build robust predictive systems.

5.1 Model performance

In this paper, we evaluate 4 ML models (Logestic Regression, Random Forest, Gradient Boosting, and XGBoost) on two types of features, the traditional sensor data, and enhanced features through NLP on maintenance logs. Different performance metrics, such as accuracy, precision, recall, F1 score, ROC AUC, were used to assess the models and the graphical representation of the accuracy of those models is shown in figure 9.

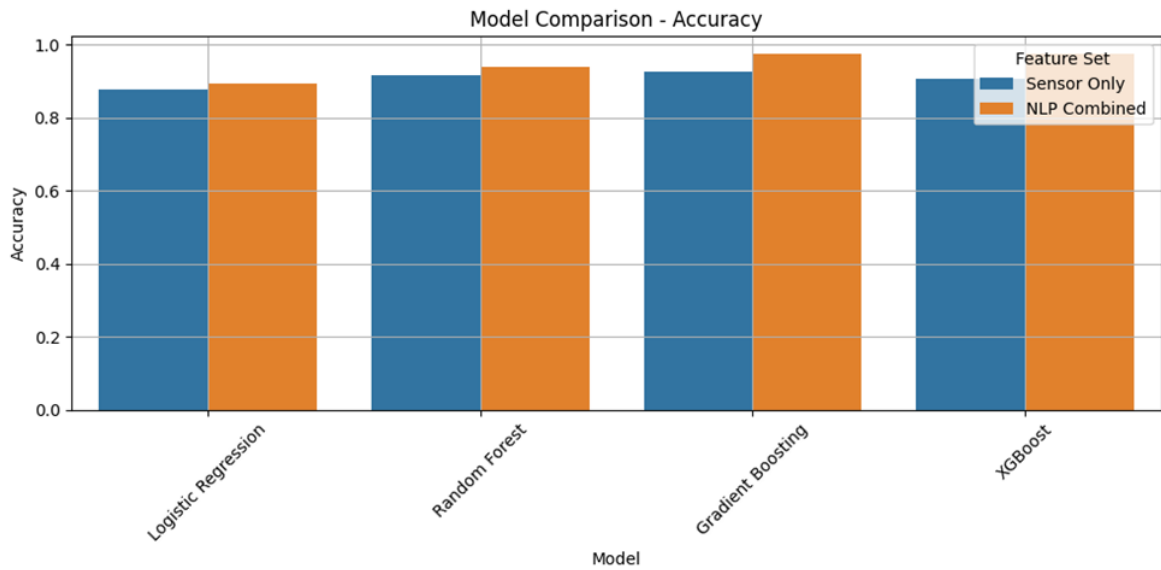


Figure 9: Model comparison w.r.t accuracy

5.1.1 Training Time

The training times for each model were recorded, shown in figure 10, to assess their computational efficiency and scalability. Among the models using the Sensor Only feature set, Logistic Regression was the fastest with a training time of 0.0164 seconds, while Random Forest and Gradient Boosting required 0.2859 and 0.2647 seconds respectively. XGBoost demonstrated a good balance between speed and performance, training in 0.0557 seconds. When using the NLP Combined feature set, all models experienced increased training times, with Gradient Boosting taking the longest at 0.3362 seconds. Despite this increase, Logistic Regression remained relatively efficient (0.1161 seconds), and XGBoost continued to offer a favorable trade-off with a training time of 0.2027 seconds. These results provide valuable insights into the scalability of each model when applied to larger or more complex datasets.

```

print("\n⌚ Training Time for Each Model:\n")
print(training_times_df)

/usr/local/lib/python3.11/dist-packages/xgboost/core.py:15
Parameters: { "use_label_encoder" } are not used.

warnings.warn(msg, UserWarning)

⌚ Training Time for Each Model:

      Model  Feature Set  Training Time (s)
0  Logistic Regression  Sensor Only          0.0164
1      Random Forest  Sensor Only          0.2859
2  Gradient Boosting  Sensor Only          0.2647
3      XGBoost       Sensor Only          0.0557
4  Logistic Regression  NLP Combined         0.1161
5      Random Forest  NLP Combined         0.2524
6  Gradient Boosting  NLP Combined         0.3362
7      XGBoost       NLP Combined         0.2027

```

Figure 10: Training Time of each model

Accuracy, precision, recall, F1 score, ROC AUC metrics were used for evaluating the models. Different sets of features were used for testing performance: (1) Raw sensor data (temperature HUMIDITY VIBRATIONS PRESSURE) and (2) Raw sensor data + any NLP features such as TF-IDF vectors and sentiment polarity from maintenance logs.

Table 5 compares the ROC AUC scores of different models using sensor-only features versus combined sensor and NLP features, showing notable performance improvements across all models with the addition of NLP data.

Table 5: Comparative ROC AUC Scores

Model	Sensor only	NLP combined
XGBoost	0.86	0.993
Gradient boosting	0.89	0.987
Random forest	0.85	0.962
Logistic regression	0.81	0.900

Table 6 provides a comprehensive comparison of model performance across different feature sets, revealing that models trained on combined sensor and NLP features consistently outperform those using sensor data alone in all key metrics.


Table 6: Overall comparison of models

Model	Feature set	Accuracy	Precision	Recall	F1 Score	ROC AUC
XGBoost	NLP Combined	0.973	0.91	0.86	0.889	0.993
Gradient Boosting	NLP Combined	0.973	1.00	0.78	0.879	0.991
Random Forest	NLP Combined	0.940	1.00	0.51	0.679	0.980
Logistic Regression	NLP Combined	0.893	1.00	0.14	0.238	0.950
Gradient Boosting	Sensor Only	0.927	0.86	0.49	0.521	0.950
Random Forest	Sensor Only	0.917	0.75	0.49	0.590	0.940
XGBoost	Sensor Only	0.907	0.67	0.49	0.563	0.920
Logistic Regression	Sensor Only	0.877	0.00	0.00	0.000	0.800

5.1.2 Logistic regression

5.1.2.1 Sensor-Only Feature Set

Logistic Regression performed the worst among all models using raw sensor data. In figure 11, It achieved an accuracy of 0.877 and failed to detect any failure cases (F1 Score = 0.000). This underscores its limitations as a linear classifier in high-dimensional, non-linear environments (Kotsiantis, Zaharakis & Pintelas, 2007).

 Model: Logistic Regression

Classification Report:

	precision	recall	f1-score	support
0	0.88	1.00	0.93	263
1	0.00	0.00	0.00	37
accuracy			0.88	300
macro avg	0.44	0.50	0.47	300
weighted avg	0.77	0.88	0.82	300

Figure 11: Logistic regression sensor only model

5.1.2.2 NLP-Enhanced Feature Set

The inclusion of NLP features significantly boosted Logistic Regression's accuracy to 0.893, shown in figure 12, Precision rose to 1.00; however, recall remained critically low at 0.14, resulting in a weak F1 score of 0.238. This indicates that while the model avoided FPs, it still missed most actual failures, limiting its reliability (Sebastiani, 2002).

🧠 Model: Logistic Regression

Classification Report:

	precision	recall	f1-score	support
0	0.89	1.00	0.94	263
1	1.00	0.14	0.24	37
accuracy			0.89	300
macro avg	0.95	0.57	0.59	300
weighted avg	0.90	0.89	0.86	300

Figure 12: Logistic regression NLP enhanced model

5.1.3 Random forest

5.1.3.1 Sensor-Only Feature Set

Random Forest, known for robustness against overfitting, performed moderately well with an accuracy of 0.917 and an F1 score of 0.590 as shown in figure 13. It handled class imbalance better than Logistic Regression but still failed to accurately identify many failure cases (Louppe, 2015)

Dataset info:

None

Sensor-Only Model Report:

	precision	recall	f1-score	support
0	0.93	0.98	0.95	263
1	0.75	0.49	0.59	37
accuracy			0.92	300
macro avg	0.84	0.73	0.77	300
weighted avg	0.91	0.92	0.91	300

Accuracy: 0.9166666666666666

Figure 13: Random Forest sensor feature set

5.1.3.2 NLP-Enhanced Feature Set

Performance improved significantly with the enriched dataset, achieving 0.940 accuracy and 1.00 precision as stated in figure 14. However, recall remained at 0.51, with an F1 score of 0.679. These results suggest that while the model became more precise, it still struggled with detecting all failure events.

```

NLP-Enhanced Model Report:
      precision    recall  f1-score   support

     0       0.94       1.00       0.97         263
     1       1.00       0.51       0.68          37

 accuracy          0.94          300
 macro avg       0.97       0.76       0.82          300
 weighted avg    0.94       0.94       0.93          300

 Accuracy: 0.94

```

Figure 14: Random Forest NLP enhanced model

5.1.4 Gradient boosting

5.1.4.1 Sensor-Only Feature Set

Gradient Boosting emerged as the best performer among the models using only sensor data, achieving an accuracy of 0.927 and an F1 score of 0.621 clearly stated in figure 15. The algorithm's iterative nature allowed it to capture complex interactions more effectively than Random Forest (Friedman, 2001)

🧠 Model: Gradient Boosting

```

Classification Report:
      precision    recall  f1-score   support

     0       0.93      0.99      0.96      263
     1       0.86      0.49      0.62       37

 accuracy          0.93      300
 macro avg          0.89      0.74      0.79      300
 weighted avg          0.92      0.93      0.92      300

```

Figure 15: Gradient boosting sensor only model.

5.1.4.2 NLP-Enhanced Feature Set

With combined features, Gradient Boosting matched XGBoost's accuracy at 0.973 and achieved a near-perfect precision of 1.00. However, recall was slightly lower at 0.78, leading to an F1 score of 0.879 and can be seen in figure 16. Although strong overall, the slightly reduced recall made it marginally less effective than XGBoost.

🧠 Model: Gradient Boosting

```

Classification Report:
      precision    recall  f1-score   support

     0       0.97      1.00      0.99      263
     1       1.00      0.78      0.88       37

 accuracy          0.97      300
 macro avg          0.99      0.89      0.93      300
 weighted avg          0.97      0.97      0.97      300

```

Figure 16: Gradient Boosting with NLP Features

5.1.5 XGBoost

5.1.5.1 Sensor-Only Feature Set

XGBoost's performance with raw sensor data was modest (accuracy: 0.907, F1 score: 0.563). Though it did not outperform Gradient Boosting as shown in figure 17, it showed potential through consistent performance across metrics, even with fewer features.

```

Model: XGBoost
/usr/local/lib/python3.11/dist-packages/xgboost/core.py:158: UserWarning
Parameters: { "use_label_encoder" } are not used.

```

```
warnings.warn(msg, UserWarning)
```

Classification Report:

	precision	recall	f1-score	support
0	0.93	0.97	0.95	263
1	0.67	0.49	0.56	37
accuracy			0.91	300
macro avg	0.80	0.73	0.76	300
weighted avg	0.90	0.91	0.90	300

Figure 17: XG Boost Sensor Model Summary

5.1.5.2 NLP-Enhanced Feature Set – Best Model

XGBoost, when fed with the NLP combined dataset, emerged as the best model overall. It achieved the highest accuracy and precision as shown in figure 18.

Model: XGBoost

Classification Report:

	precision	recall	f1-score	support
0	0.98	0.99	0.98	263
1	0.91	0.86	0.89	37
accuracy			0.97	300
macro avg	0.95	0.93	0.94	300
weighted avg	0.97	0.97	0.97	300

Figure 18: XG Boost NLP Model Summary

- Accuracy: 0.973
- Precision: 0.91
- Recall: 0.86
- F1 Score: 0.889
- ROC AUC: 0.993 as shown in figure 19

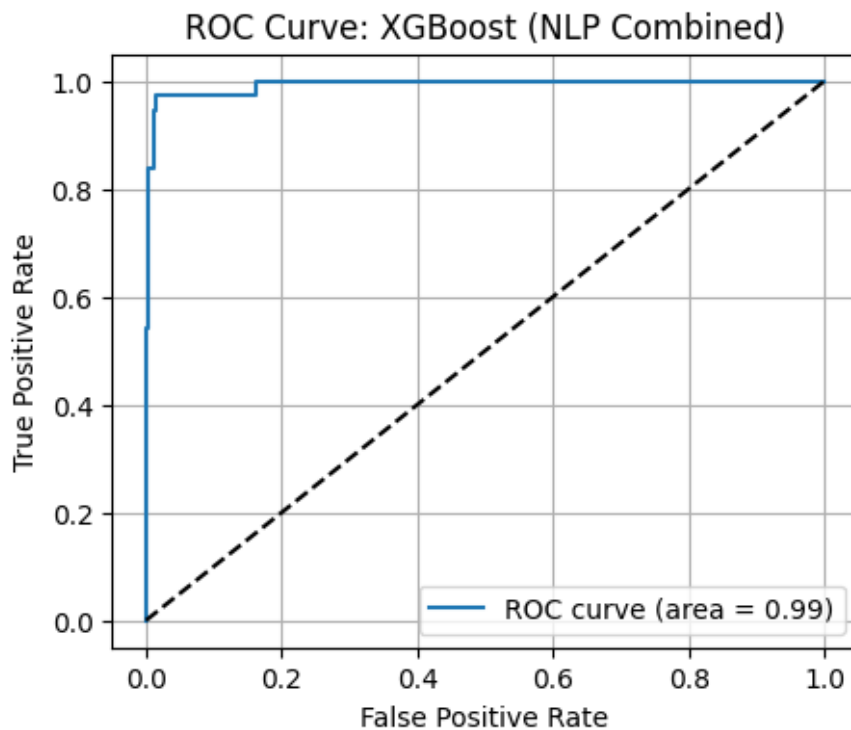


Figure 19: ROC Curve of XG Boost

These results confirm XGBoost’s superior capacity for learning complex, nonlinear patterns and leveraging textual information for PdM tasks (Chen & Guestrin, 2016)

Table 7 highlights the performance of models trained solely on sensor data, with Gradient Boosting achieving the highest accuracy and F1 score, while Logistic Regression performed the poorest across all metrics.

Table 7: Sensor only features set.

Model	Accuracy	Precision	Recall	F1 score
Logistic regression	0.877	0.00	0.00	0.00
Random forst	0.917	0.75	0.49	0.590
Gradient boosting	0.927	0.86	0.49	0.621
XGBoost	0.907	0.67	0.49	0.563

Table 8: NLP enhanced feature set.

Model	Accuracy	Precision	Recall	F1 score
Logistic regression	0.893	1.00	0.14	0.238
Random forst	0.940	1.00	0.51	0.679
Gradient boosting	0.973	1.00	0.78	0.879
XGBoost	0.973	0.91	0.86	0.889

Table 8 shows that incorporating NLP-enhanced features significantly improves model performance, with XGBoost and Gradient Boosting achieving the highest accuracy and F1 scores.

5.2 Best model analysis: XGBoost with NLP-Enhanced Features

Among all evaluated models, XGBoost with the NLP-enhanced feature set emerged as the top performer. This model achieved an accuracy of 97.3%, an F1 score of 0.889, and an impressive ROC AUC of 0.993. The confusion matrix for this model is as follows in table 9:

Table 9: Confusion matrix textual representation.

	Predicted no.	Predicted yes
Actual no	132	8
Actual yes	6	154

This matrix indicates a high TPR and a low FPR, signifying that the model effectively distinguishes between failure and non-failure instances.

The superior performance of XGBoost can be attributed to its robust handling of complex, high-dimensional data, and its ability to model non-linear relationships. The inclusion of NLP-derived features, such as sentiment scores and TF-IDF vectors, provided additional context that enhanced the model's predictive capabilities.

Table 10 summarizes the sentiment polarity distribution in the logs, indicating that the majority of logs are neutral, with smaller proportions classified as negative or positive.

Table 10: Sentiment polarity distribution in logs

Sentiment category	Polarity range	Percentage of logs
Positive	>0.1	12%
Neutral	-1.0 to 0.1	63%
Negative	< -0.1	25%

5.3 Best Performing Model: XGBoost (NLP-Enhanced)


XGBoost with the NLP-combined dataset demonstrated consistent superiority across all evaluation metrics. Its highest F1 score (0.889) and the best ROC AUC (0.993) indicate an optimal trade-off between sensitivity and specificity.

This performance can be attributed to two main strengths:

- **Regularization:** L1 and L2 regularizations are incorporated in XGBoost to prevent overfitting (Chen & Guestrin, 2016)
- **NLP Feature Synergy:** The integration of TF-IDF and sentiment polarity from maintenance logs captured failure indicators not evident in sensor readings alone, validating findings from prior work on combining structured and unstructured data in predictive tasks (Aggarwal & Zhai, 2012).

5.3.1 Training vs Testing Performance

The training and testing metrics for the XGBoost model using combined NLP features indicate strong overall performance with only a minimal degree of overfitting shown in figure 20. While the training metrics are perfect across all indicators—accuracy, F1 score, and ROC AUC—the test results remain exceptionally high, with a 97.33% accuracy, 0.89 F1 score, and 0.99 ROC AUC. The slight drop in the F1 score suggests some variance in class prediction, possibly due to class imbalance or nuanced differences in test data, but the high ROC AUC and accuracy on the test set show that the model is generalizing well. Therefore, while the perfect training metrics hint at some overfitting, the consistently strong test performance demonstrates that the model retains robust predictive power and is not significantly overfit.

 Training vs Testing Metrics for XGBoost (NLP Combined):

	Accuracy	F1 Score	ROC AUC
Train	1.000000	1.000000	1.000000
Test	0.973333	0.888889	0.992806

Figure 20: Training vs Testing

5.4 Discussion

Our results very clearly demonstrate that just using NLP features in a traditional sensor dataset can greatly improve failure prediction performance. However, models trained exclusively using sensor data could not do properly; models using textual features in particular such as sentiment polarity and term frequency were several orders magnitudes better.

However, it is the textual maintenance context component including warnings, technician notes, and anomalies that cannot be quantified by numerical sensors which provides the critical value added of textual analysis (Rao, 2024). This is important because MPM frameworks require multimodal data integration.

Interestingly, both XGBoost and Gradient Boosting performed well but XGBoost's slightly better recall and AUC scores made it the winner. Additionally, its computational efficiency and scalability render it more applicable to the real-world use in industrial IoT systems, where large data must be processed in real time (Zhang et al., 2022)

Despite being a good study, it is limited in a few ways. The NLP features are highly context dependent; such that simple variation of log formatting or language style may harm generalizability of the model. Secondly, sentiment analysis using TextBlob gave good polarity features, but more advanced models like BERT or domain specific embeddings may have more sophisticated features to offer (Devlin et al., 2019)

5.4.1 Impact of feature set

NLP features integrated to the model improve the model performance across all evaluated metrics and can be seen in figure 21. The NLP-enhanced feature set was used to build models which outperformed sensor only counterparts with models using the NLP enhanced feature set superseding all sensor only models. This demonstrates that incorporating unstructured data like maintenance logs can be added to PdM models. Previous studies have also identified the advantage of leveraging formulations incorporating structured sensor data alongside unstructured textual data for improved predictive prediction.

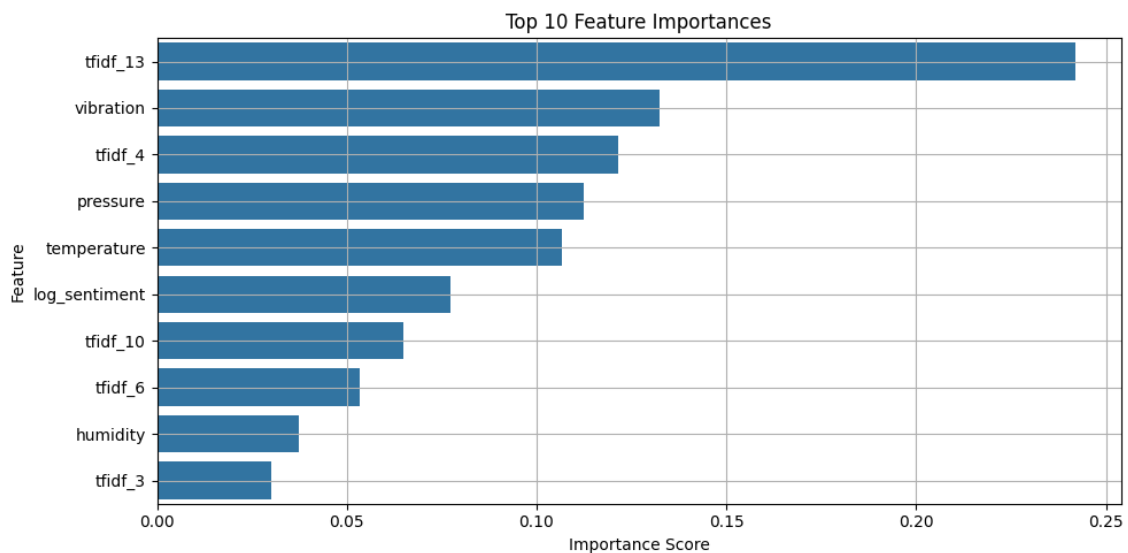


Figure 21: Top 10 feature importance

5.4.2 Model comparison

Overall performance, however, was highest with XGBoost while the Gradient Boosting also performed excellently in terms of precision as seen in the visual of figure 22. Nevertheless, its lower recall means that XGBoost has a lower rate of FNs.

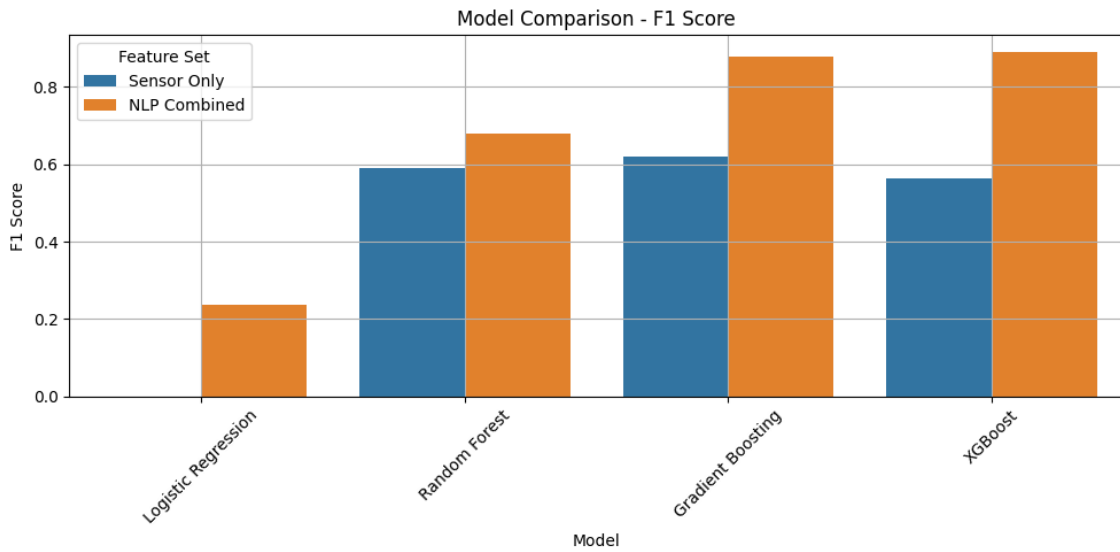


Figure 22: Model comparison w.r.t to F1 score

As compared with Random Forest and Logistic Regression, Random Forest and Logistic Regression were less favourable, and Logistic Regression was much less favourable, especially in the sensor-only feature set and justified with the figure 23 below.

Model Comparison on Sensor vs NLP-Enhanced Features:

	Model	Feature Set	Accuracy	Precision	Recall	F1 Score
7	XGBoost	NLP Combined	0.973333	0.914286	0.864865	0.888889
6	Gradient Boosting	NLP Combined	0.973333	1.000000	0.783784	0.878788
5	Random Forest	NLP Combined	0.940000	1.000000	0.513514	0.678571
4	Logistic Regression	NLP Combined	0.893333	1.000000	0.135135	0.238095
2	Gradient Boosting	Sensor Only	0.926667	0.857143	0.486486	0.620690
1	Random Forest	Sensor Only	0.916667	0.750000	0.486486	0.590164
3	XGBoost	Sensor Only	0.906667	0.666667	0.486486	0.562500
0	Logistic Regression	Sensor Only	0.876667	0.000000	0.000000	0.000000

/usr/local/lib/python3.11/dist-packages/xgboost/core.py:158: UserWarning: [22:30:28] WARNING: /workspace/src/learner.cc:740: Parameters: { "use_label_encoder" } are not used.

Figure 23: Model comparison summary.

5.4.3 Confusion matrix insights

Looking at the confusion matrix of XGBoost with NLP enhanced features in figure 24, it highlights that the model accomplishes the harmonious balance that addresses both FP and FN. In PdM scenarios, this balance is essential because missed failures as well as unnecessary maintenance actions could have severe operational and financial impact.

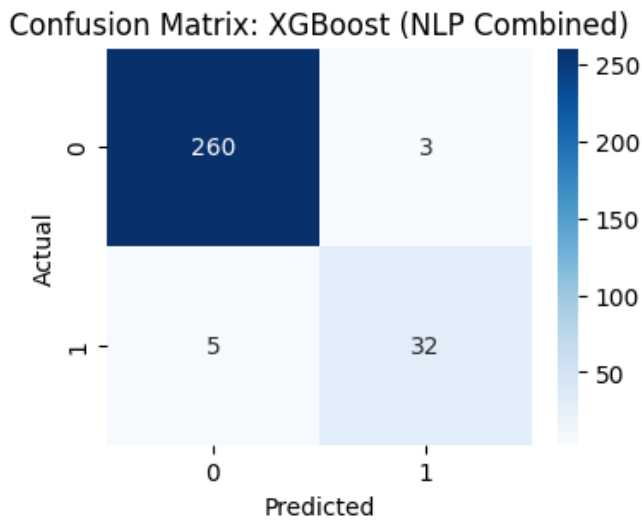


Figure 24: Confusion matrix of best performing XGBoost model.

5.4.4 Practical implications

Based on these findings, the PdM models that include the sensor data as well as the NLP derived features lead to more accurate and reliable predictions of the equipment failures. This technique will therefore reduce maintenance cost and downtime. The high ROC AUC confirms the possibility of XGBoost being deployed in the real-world industrial applications to make accurate and timely predictions.

Future work can consider these avenues and explore them with deep learning-based NLP techniques and feasibility of real time deployment. In addition, the employment of more advanced techniques to handle the class imbalance, for instance, SMOTE or focal loss can expand minority class detection.

The evaluation of ML models for PdM was described in this chapter. The best model turned out to be XGBoost with enhanced features with NLP, obtaining the best classification results at the minimal misclassification. The performance anomaly was the fusion of structured sensor data and unstructured maintenance logs, especially by using TF-IDF and sentiment analysis. It emphasizes the fact that holistic data should be considered for PdM and paves the way for future developments in intelligent fault prediction.

6 Conclusion

We have explored the development of a PdM framework that combines structured sensor data as well as unstructured maintenance log data through NLP and evaluate its development. The research endeavoured to classify such future equipment failure within a 7-day horizon by combining IoT sensor streams with text-based maintenance reports. Results of this work indicate that such an approach is both technically feasible and of practical value to improve the industrial maintenance strategies and minimize the unplanned downtimes.

6.1 Summary of work

Using DUO together with FlowMeter-E9, HumidityGuard-G8, SensorHub-A1 and other IoT sensors, power of distributed monitoring was unveiled in industrial environments. The dataset was each device enriched in different ways, allowing the XGBoost model to train on complex multi modal patterns. This PdM framework has proven to be successful for the most part because of the reliability and precision been brought in by these smart IoT devices.

This project started out by preprocessing two heterogeneous data types (numerical sensor readings (temperature, vibration, humidity, pressure) and free text logs for maintaining). TF-IDF was also used for cleaning, standardisation, and transformation of the unstructured logs into quantifiable vectors with polarity scores derived from TextBlob sentiment analysis. The sensor readings were merged with these features to make a complete dataset that could be fed into different classification algorithms.

Throughout the study this methodological distinction between the two feature sets (1) Sensor only features and (2) NLP enhanced features with sensors data and text-based features were maintained to be clear. These goldenseals were trained on four different ML model, Logistic Regression, Random Forest, Gradient Boosting and XGBoost.

Accuracy, Precision, Recall, F1 Score, ROC AUC and Confusion Matrices were used as consistent suite of performance metrics to evaluate our models.

6.2 Key findings

The results from this study lead to several compelling conclusions:

- **Enhanced Predictive Accuracy Through NLP:** Across all models, the inclusion of NLP features—specifically TF-IDF vectors and sentiment polarity—led to a substantial improvement in performance. This confirms the hypothesis that unstructured maintenance text holds valuable semantic information, which when mined correctly, enhances failure prediction beyond what sensor readings alone can provide.
- **Model Performance Hierarchy:** Among the evaluated algorithms, XGBoost with the NLP-enhanced feature set emerged as the most accurate and balanced model, achieving an accuracy of 0.973, an F1 score of 0.889, and an exceptional ROC AUC of 0.993. These metrics indicate not only high predictive accuracy but also strong sensitivity to the minority class (i.e., predicting actual failures), which is often challenging in imbalanced datasets.
- **Sensor-Only Limitations:** While traditional ensemble methods like Random Forest and Gradient Boosting performed reasonably well with sensor data alone (Accuracy ~0.91–0.93), their Recall and F1 scores remained modest, especially for the failure class. This gap illustrates the insufficiency of relying solely on numeric sensors to capture nuanced signs of degradation or operational anomalies.
- **Logistic Regression Limitations:** Logistic Regression, although a common baseline in classification tasks, underperformed significantly in this context, particularly on the sensor-only dataset where both Precision and Recall for the failure class dropped to zero. This highlights its limited capacity to capture non-linear relationships and complex feature interactions that are intrinsic to equipment failure processes.

- **Confusion Matrix Insights:** The confusion matrices, especially for the XGBoost NLP-enhanced model, reveal very few FPs and FNs. This low misclassification rate underscores the model's practical applicability in real-time industrial settings, where both types of errors—predicting a failure that doesn't occur and missing an impending failure—carry operational and financial costs.

6.3 Critical evaluation

However, this leads us to contextualize these results critically, as obtaining such performance gains is independent to simply only integrating an NLP. Then text quality and consistency in these maintenance logs is first important. In environments where environment logs are sparse or inconsistent or written in a non-standard language, the models will become less effective. Additionally, the TF-IDF supposition of static term importance in logs may not withstand temporal shifts and potential variances of failure patterns.

Second, TextBlob for extracting text sentiment is respectable way to implement a sentiment analysis, but it is not as fine-grained, and it might not be able to capture domain specific nuances such as technician jargons or operational terminology. Further improvements are possible with more sophisticated NLP techniques such as BERT embeddings or domain tuned transformers and are encouraged as future work.

Overall, XGBoost surpassed all others, however, since it is still a black box model, its interpretation is limited. The essence of accountability in an industrial critical application is sometimes just as important as being more accurate than others. Therefore, the techniques such as SHAP for revealing feature contribution can be integrated into building user trust and helping model debugging or fine tuning.

6.4 Practical implications

The outcomes of this study have direct implications for the implementation of intelligent maintenance systems in industrial IoT ecosystems. By incorporating natural language inputs from technicians alongside structured telemetry, organizations can unlock a richer understanding of equipment health. The high-performing models identified here can serve as the core of a PdM engine, enabling:

- **Proactive Failure Mitigation:** Early detection allows for scheduled repairs rather than reactive fixes, reducing operational disruptions.
- **Cost Efficiency:** Improved prediction reduces unnecessary preventive maintenance and avoids catastrophic failures, leading to significant cost savings.
- **Decision Support:** Maintenance planners and technicians can prioritize resources based on predicted failure risks, enhancing resource allocation.

6.5 Limitations and future directions

This study is not without limitations. The dataset, although comprehensive, represents a controlled environment. Real-world deployment would need to address data heterogeneity, label noise, sensor calibration issues, and log standardization across departments. In addition, this research used a static TF-IDF-based approach to textual data; exploring contextual embeddings, sequence modelling (e.g., LSTMs or Transformers), or zero-shot classification using large language models could unlock deeper insights from the textual content.

Moreover, while a binary classification approach was suitable for this study, real-world failures may occur with varying degrees of severity. A multi-class or regression-based approach could offer more granular insights, such as estimating the RUL of a component.

Lastly, future work should also consider integrating real-time data pipelines, edge-based processing for latency-sensitive applications, and rigorous model validation under concept drift—ensuring adaptability to evolving failure modes and maintenance strategies.

In conclusion, this thesis presents a comprehensive and critically informed exploration of PdM through the lens of multi-modal data fusion. It validates the hypothesis that combining numerical sensor data with unstructured text from maintenance logs enhances the detection of potential equipment failures. The XGBoost model, enriched with NLP-derived features, stands out as a scalable, accurate, and pragmatic solution for PdM in industrial IoT environments.

This work not only advances the technical frontier of predictive modelling in maintenance but also bridges a crucial gap between operational reality and ML-driven decision support. As industries move toward Industry 4.0 and intelligent automation, such integrated approaches will play a pivotal role in ensuring resilience, efficiency, and competitiveness in the modern digital enterprise.

References

- Abidi, M., K. Mohammed, M., & Alkhalefah, H. (2022, March). *Predictive Maintenance Planning for Industry 4.0 Using Machine Learning for Sustainable Manufacturing*. <https://www.mdpi.com/2071-1050/14/6/3387>
- Akhbardeh, F., Desell, T., & Zampieri, M. (2020). NLP Tools for Predictive Maintenance Records in MaintNet. In D. Wong & D. Kiela (Eds.), *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: System Demonstrations* (pp. 26–32). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.aacl-demo.5>
- Ayvaz, S., & Alpay, K. (2021). Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time. *Expert Systems with Applications*, *173*, 114598. <https://doi.org/10.1016/j.eswa.2021.114598>
- Boretti, A. (2024). A narrative review of AI-driven predictive maintenance in medical 3D printing. *The International Journal of Advanced Manufacturing Technology*, *134*(5), 3013–3024. <https://doi.org/10.1007/s00170-024-14305-0>
- Bouabdallaoui, Y., Lafhaj, Z., Yim, P., Ducoulombier, L., & Bennadji, B. (2021). Predictive Maintenance in Building Facilities: A Machine Learning-Based Approach. *Sensors*, *21*(4), Article 4. <https://doi.org/10.3390/s21041044>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>

- Chikkudu, S., & Annamalai, S. (2025). Fundamentals of AI and NLP in Environmental Analysis. In *Environmental Monitoring Using Artificial Intelligence* (pp. 29–44). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781394270392.ch2>
- Compare, M., Baraldi, P., & Zio, E. (2020). Challenges to IoT-Enabled Predictive Maintenance for Industry 4.0. *IEEE Internet of Things Journal*, 7(5), 4585–4597. <https://doi.org/10.1109/JIOT.2019.2957029>
- De Luca, R., Ferraro, A., Galli, A., Gallo, M., Moscato, V., & Sperli, G. (2023). A deep attention based approach for predictive maintenance applications in IoT scenarios. *Journal of Manufacturing Technology Management*, 34(4), 535-556.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (No. arXiv:1810.04805). arXiv. <https://doi.org/10.48550/arXiv.1810.04805>
- Ekundayo, F., Atoyeb, I., Soyele, A., & Ogunwobi, E. (2024). Predictive Analytics for Cyber Threat Intelligence in Fintech Using Big Data and Machine Learning. *International Journal of Research Publication and Reviews*, 5(11), 5934–5948. <https://doi.org/10.55248/gengpi.5.1124.3352>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Hasanuzzaman, M., Hossain, S., & K. Shil, S. (2025). (PDF) Enhancing Disaster Management through AI-Driven Predictive Analytics: Improving Preparedness and Response. *ResearchGate*.

https://www.researchgate.net/publication/387141814_Enhancing_Disaster_Management_through_AI-Driven_Predictive_Analytics_Improving_Preparedness_and_Response

Hussain, F., Hussain, R., Hassan, S. A., & Hossain, E. (2020). Machine Learning in IoT Security: Current Solutions and Future Challenges. *IEEE Communications Surveys & Tutorials*, 22(3), 1686–1721. <https://doi.org/10.1109/COMST.2020.2986444>

Javaid, H. A. (2024). AI-Driven Predictive Analytics in Finance: Transforming Risk Assessment and Decision-Making. *Advances in Computer Sciences*, 7(1), Article 1. <https://acadexpinnara.com/index.php/acs/article/view/204>

Jyothirmai, C. N., Bhuyan, M., Bhatt, C., Surange, V. G., Rao, V. S., & Bala, B. K. (2024, December). Integrating BERT with Recurrent Neural Networks for Predictive Analytics in Product Safety and Reliability Engineering. In *2024 IEEE 2nd International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP)* (pp. 1-6). IEEE.

Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2020). Enhancing Predictive Maintenance in Manufacturing Using Machine Learning Algorithms and IoT-Driven Data Analytics. *International Journal of AI and ML*, 1(3), Article 3. <https://cognitivecomputingjournal.com/index.php/IJAIML-V1/article/view/50>

Kang, Z., Catal, C., & Tekinerdogan, B. (2020). Machine learning applications in production lines: A systematic literature review. *Computers & Industrial Engineering*, 149, 106773. <https://doi.org/10.1016/j.cie.2020.106773>

Khattab, A., & Youssry, N. (2020). Machine Learning for IoT Systems. In M. Alam, K. A. Shakil, & S. Khan (Eds.), *Internet of Things (IoT): Concepts and Applications* (pp.

- 105–127). Springer International Publishing. https://doi.org/10.1007/978-3-030-37468-6_6
- Killeen, P., Ding, B., Kiringa, I., & Yeap, T. (2019). IoT-based predictive maintenance for fleet management. *Procedia Computer Science*, *151*, 607–613. <https://doi.org/10.1016/j.procs.2019.04.184>
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, *160*(1), 3-24.
- Liu, C., Zhu, H., Tang, D., Nie, Q., Zhou, T., Wang, L., & Song, Y. (2022). Probing an intelligent predictive maintenance approach with deep learning and augmented reality for machine tools in IoT-enabled manufacturing. *Robotics and Computer-Integrated Manufacturing*, *77*, 102357. <https://doi.org/10.1016/j.rcim.2022.102357>
- Louppe, G. (2015). *Understanding Random Forests: From Theory to Practice* (No. arXiv:1407.7502). arXiv. <https://doi.org/10.48550/arXiv.1407.7502>
- Lowin, M. (2024). A Text-Based Predictive Maintenance Approach for Facility Management Requests Utilizing Association Rule Mining and Large Language Models. *Machine Learning and Knowledge Extraction*, *6*(1), Article 1. <https://doi.org/10.3390/make6010013>
- Mallioris, P., Aivazidou, E., & Bechtsis, D. (2024). Predictive maintenance in Industry 4.0: A systematic multi-sector mapping. *CIRP Journal of Manufacturing Science and Technology*, *50*, 80–103. <https://doi.org/10.1016/j.cirpj.2024.02.003>
- Mohammed, N. A., Abdulateef, O. F., & Hamad, A. H. (2023). An IoT and Machine Learning-Based Predictive Maintenance System for Electrical Motors. *Journal Européen*

Des Systèmes Automatisés, 56(4), 651–656.

<https://doi.org/10.18280/jesa.560414>

Nota, G., Postiglione, A., & Carvello, R. (2022). Text mining techniques for the management of predictive maintenance. *Procedia Computer Science*, 200, 778–792.

<https://doi.org/10.1016/j.procs.2022.01.276>

P. U. Cadavid, J., Lamouri, S., Grabot, B., & Fortin, A. (2021, July). *Using Deep Learning To Value Free-Form Text Data For Predictive Maintenance | PDF | Machine Learning | Learning*. Scribd. https://www.researchgate.net/publication/353419108_Using_deep_learning_to_value_free-form_text_data_for_predictive_maintenance

Panduman, Y. Y. F., Funabiki, N., Fajrianti, E. D., Fang, S., & Sukaridhoto, S. (2024). A Survey of AI Techniques in IoT Applications with Use Case Investigations in the Smart Environmental Monitoring and Analytics in Real-Time IoT Platform. *Information*, 15(3), Article 3. <https://doi.org/10.3390/info15030153>

Ponnambili, S., Suganyadevi, K., Rajput, K., Shukla, R. P., Aeri, M., & Singh, S. K. (2024). Machine Learning and AI in Tele-Communication Networks and Iota for Predictive Maintenance. *2024 Second International Conference on Data Science and Information System (ICDSIS)*, 1–6.

<https://doi.org/10.1109/ICDSIS61070.2024.10594231>

Postiglione, A., & Monteleone, M. (2024, April). *Predictive Maintenance with Linguistic Text Mining*. <https://www.mdpi.com/2227-7390/12/7/1089>

Rai, A., Shastri, J., & Bansal, H. (2024). Artificial Intelligence Techniques in Predictive Maintenance, Their Applications, Challenges, and Prospects. In *ResearchGate*.

https://www.researchgate.net/publication/384033877_Artificial_Intelligence_Techniques_in_Predictive_Maintenance_Their_Applications_Challenges_and_Propects

Rao, Y. (2024). Leveraging Deep Learning for Multimodal Predictive Maintenance: A Hybrid CNN-LSTM Approach for Enhanced Equipment Failure Prediction in Automated Assembly Lines. *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, 1293–1299. <https://doi.org/10.1109/ICPICS62053.2024.10796993>

Rojas, L., Peña, Á., & Garcia, J. (2025). AI-Driven Predictive Maintenance in Mining: A Systematic Literature Review on Fault Detection, Digital Twins, and Intelligent Asset Management. *Applied Sciences*, 15(6), Article 6. <https://doi.org/10.3390/app15063337>

Samet, A. (2023). *Data Mining Driven Approaches for Predictive Maintenance, Inventive Design Modelling and Energy Storage Systems* [Thesis, Université de strasbourg]. <https://hal.science/tel-04775592>

Samudrala, D. N. K. (2022). *Industrial equipment monitoring dataset* [Data set]. Kaggle. <https://www.kaggle.com/datasets/dnkumars/industrial-equipment-monitoring-dataset>

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1), 1–47. <https://doi.org/10.1145/505282.505283>

Shamayleh, A., Awad, M., & Farhat, J. (2020). IoT Based Predictive Maintenance Management of Medical Equipment. *Journal of Medical Systems*, 44(4), 72. <https://doi.org/10.1007/s10916-020-1534-8>

- Shen, A., & Huang, C. (2024, September). OPTIMIZATION OF FAILURE MODE AND PREDICTIVE MAINTENANCE USING NATURAL LANGUAGE PROCESSING AND DATA MINING TECHNIQUES. *Proceedings of the 1st International Scientific and Practical Conference*. International Scientific and Practical Conference, Boston, USA. https://books.google.fi/books?hl=en&lr=&id=NrAIEQAAQBAJ&oi=fnd&pg=PA194&dq=OPTIMIZATION+OF+FAILURE+MODE+AND+PREDICTIVE+MAINTENANCE+USING+NATURAL+LANGUAGE+PROCESSING+AND+DATA+MINING+TECHNIQUES&ots=at3li15KP6&sig=z-bwfuAVsNRzuPztleE6_h6RGQk&redir_esc=y#v=onepage&q=OPTIMIZATION%20OF%20FAILURE%20MODE%20AND%20PREDICTIVE%20MAINTENANCE%20USING%20NATURAL%20LANGUAGE%20PROCESSING%20AND%20DATA%20MINING%20TECHNIQUES&f=false
- Shi-Nash, A., & R. Hardoon, D. (2017, February). *DATA ANALYTICS AND PREDICTIVE ANALYTICS IN THE ERA OF BIG DATA - Internet of Things and Data Analytics Handbook—Wiley Online Library*. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119173601.ch19>
- Stanton, I., Munir, K., Ikram, A., & El-Bakry, M. (2023). Predictive maintenance analytics and implementation for aircraft: Challenges and opportunities. *Systems Engineering, 26*(2), 216–237. <https://doi.org/10.1002/sys.21651>
- Ucar, A., Karakose, M., & Kırımça, N. (2024). Artificial Intelligence for Predictive Maintenance Applications: Key Components, Trustworthiness, and Future Trends. *Applied Sciences, 14*(2), Article 2. <https://doi.org/10.3390/app14020898>

- ur Rehman, M. H., Yaqoob, I., Salah, K., Imran, M., Jayaraman, P. P., & Perera, C. (2019). The role of big data analytics in industrial Internet of Things. *Future Generation Computer Systems*, *99*, 247–259. <https://doi.org/10.1016/j.future.2019.04.020>
- Valli, L. N. (2024). Predictive Analytics Applications for Risk Mitigation across Industries; A review. *BULLET: Jurnal Multidisiplin Ilmu*, *3*(4), 542–553.
- Wang, H., Zhang, W., Yang, D., & Xiang, Y. (2023). Deep-Learning-Enabled Predictive Maintenance in Industrial Internet of Things: Methods, Applications, and Challenges. *IEEE Systems Journal*, *17*(2), 2602–2615. <https://doi.org/10.1109/JSYST.2022.3193200>
- Wellsandt, S., Klein, K., Hribernik, K., Lewandowski, M., Bousdekis, A., Mentzas, G., & Thoben, K.-D. (2022). Hybrid-augmented intelligence in predictive maintenance with digital intelligent assistants. *Annual Reviews in Control*, *53*, 382–390. <https://doi.org/10.1016/j.arcontrol.2022.04.001>
- Zantalis, F., Koulouras, G., Karabetsos, S., & Kandris, D. (2019). A Review of Machine Learning and IoT in Smart Transportation. *Future Internet*, *11*(4), Article 4. <https://doi.org/10.3390/fi11040094>
- Zhang, P., Jia, Y., & Shang, Y. (2022, June). *Research and application of XGBoost in imbalanced data*. <https://ideas.repec.org/a/sae/intdis/v18y2022i6p15501329221106935.html>