**CCT College Dublin**

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | **Advance Data Analytics**  **Big Data Storage and Processing** |
| **Assessment Title:** | **CA 1 Sem 2 MSc in Data Analytics** |
| **Lecturer Name:** | **David McQuaid, Muhammad Iqbal** |
| **Student Full Name:** | **Muhammad Asif** |
| **Student Number:** | **2021402** |
| **Assessment Due Date:** | **07-4-2024** |
| **Date of Submission:** | **7-4-2024** |

**Muhammad asif 7-4-2024**

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

# Predictive Maintenance Using Long Short-Term Memory (LSTM) Networks

# Abstract

This project concentrates on the implementation of various analytical approaches, e.g. neural networks, into storage and processing data for the predictive maintenance. The mentioned study investigates solutions in the approach of machine learning techniques such as LSTM networks and distributed computing platform like Apache Spark to conclude that predictive maintenance models can be trusted and highly efficient in the forecasting of equipment failures. The main research conclusions are the universality of distributed computing architectures that make them ready for the processing of the whole set of sensor data and the possibility to use preventive maintenance systems created for industrial automation. This exercise is a move on that proves to be both informative and useful by providing data analytics capabilities in an industrial setting.

***Keywords****: Predictive Maintenance, Neural Networks, Big Data Analytics, Distributed Computing*

# INTRODUCTION

## *Background and Motivation*

Now, in the industrial sector, predictive maintenance is considered the main tool to increase productivity and to reduce downtime, and as a result to save the resource allocation costs. The traditional maintenance scheduling process is a combination of pre-determined coverages or unexpected repairs which result into inefficiency and higher costs. It is novel that new data analytics tools along with the rise of the spread of the sensor technologies enable enterprises to change to data-driven and predictive maintenance strategies.

The idea for the project was inspired by the general notion that preventive maintenance plays a great role in the industrial process’s transformation. We are going to design a predictive maintenance system which will be able to anticipate equipment failures in advance helps to utilize the newest machine learning methods such as LSTM (Long Short-Term Memory) network among the others. On the other hand, this is a preventive approach that results in less unforeseen downtime and offers space to the businesses to tailor their maintenance schedule, as well as to improve operations, and extend the life of vital equipment.

Also, the integration of data analytics, big data storage and processing brings in newly emerged possibilities and problems. Industrial sensors are incredibly source to the data that is fast flowing and in the wide range of volumes, which raises the need for recording systems that are scalable and affordable and the analytics systems that can uncover the useful meanings from the huge data sets.

## *Problem Statement*

The traditional strategy is what most organizations use which involves exploring the equipment on a regular basis and reactively resolving on equipment faults. It usually ends up with the loose of a time, and inefficient use of resources, and moreover this leads to the decreasing in the operation output. The main issue here is that we cannot predict and thus avoid such accidents, resulting in unplanned downtimes and a consequent production loss.

* ***Unplanned Downtime:*** *The lack of predictive maintenance skills exposes industrial firms to unplanned downtime, which can have serious financial consequences and interrupt production plans.*
* ***Suboptimal Resource Allocation:*** *Traditional maintenance approaches frequently result in suboptimal resource allocation, with maintenance operations occurring either too frequently, incurring needless expenses, or seldom, increasing the risk of failure.*
* ***Reactive Maintenance:*** *Reactive solutions to equipment failures not only increase maintenance costs, but they also fail to address underlying issues, resulting in recurring failures and shorter equipment lifespans.*
* ***Limited Data Utilization:*** *Despite the quantity of sensor data generated by industrial equipment, many businesses are unable to fully utilize this data for predictive maintenance owing to restrictions in data storage, processing, and analytics.*

This project aims to overcome the stated problems by designing a maintenance prediction system that is built on LSTM networks, which fall under the category of RNN and that can process historical data from the sensors and foretell equipment failure. The first objective of the methodology is to wipe out downtime, facilitate resource sharing, and enhance operational reliability by foreseeing maintenance needs and setting up repairs before failures occur.

In addition, the program would be analyzing the consequences of managing large sensor Data in a Big Data storage and processing environment. The purpose is to deal with the scalability issue and to extract the critical information from massive databases which will contribute to the improvement of more efficient monitoring models.

## *Research Question*

To help this initiative achieve its goals, the following research question is posed:

*"How can the LSTM networks be efficiently used to predict the manufacturing systems maintenance and what place do the Big Data storage and processing technologies have for this purpose?”*

This research question includes many major factors that will be investigated during the project:

* *This project is designed to investigate which type of LSTM networks can be used for the prediction of equipment degradation and maintenance required by the historical sensor data. This includes assessing the trustworthiness, precision, and ability to scale in the actual circumstances where the LSTM-based preventive maintenance model is being deployed.*
* *In this study, it is targeted to test the integration of predictive maintenance models based on LSTM with the big data storage and processing tech. This is done by identifying and resolving data sensory issues, storing, and retrieving data storage methods more effectively, and using advanced analytical tools to take insights from big data.*
* *The research will be focusing on the practical/tangible effects of the application of predictive maintenance systems (based on LTSM) in the industrial sector. Such a system can be very complicated and difficult to implement, it will require to be deployed and synchronized with current maintenance, and the practice and procedures.*

# LITERATURE REVIEW

## *Introduction to Predictive Maintenance*

Predictive maintenance in its essence is a proactive maintenance approach, which forecasts and prevent the breakdown of equipment before failure occurs, which in turn reduces the downtime, decreases the maintenance budget, and increases the reliability of assets. The predictive maintenance method is quite different from the old-school timetable or reactive maintenance as the latter are only reactive and fix the equipment failures. Predictive maintenance uses modern data analytics techniques along with sensor technologies which can predict the equipment health and maintenance requirements using real-time or historical data.

Predictive maintenance has gained a lot of interest among many industries like manufacturing, energy, transport, and healthcare since there is an increase in sensor data nowadays and the mathematical algorithms and models for machine learning are advancing as well as the urge for achieving more efficient operations and asset utilization [2]. As the organization moves from time-based reactive maintenance to predictive maintenance, it will be able to increase the equipment performance, improve safety standards and reduce downtime.

Predictive maintenance is founded on models and algorithms, which are used to predict machine failures by processing sensor data. These models could be used to spot abnormalities, predict trends, which may indicate impending failures, and the maintenance work recommendations [4]. The LSTM networks have been found to perform well as an efficient tool for predicting multi-modal and complex time-series data [5].

The major features of predictive maintenance can be summarized as the following:

* ***Reduced Downtime:*** *Predictive maintenance will be done by detecting and fixing the potential failures and interruptions before they can happen, so that the occurrence of failures and downtime, which may lead to slowing down the operations, will be reduced, thus ensuring reliable operations and increased productivity [6].*
* ***Cost Savings:*** *Predictive maintenance creates an avenue for firms to optimize maintenance plans, reduce the unexpected contingency costs and prolong the useful life of assets.*
* ***Improved Asset dependability:*** *Predictive maintenance, which uses preventive measures like equipment monitoring and repairs at the right time, is the most effective way to extend the life and reliability of machinery while avoiding the risk of catastrophic failure.*
* ***Enhanced Safety:*** *Preventive maintenance helps identify the safety problems that may be due to equipment incapacity and hence reduces the risks of the work [8].*

## Overview of Long Short-Term Memory (LSTM) Networks

LSTM is an RNN-based model that can establish the relationship between things such as images, words, and input sequences over the long-term. Unlike other RNNs which generate the vanishing gradient problem and have severe memory loss on long time sequences, LSTM networks employ memory cells and gates and process information efficiently over long time periods [10].

The LSTM network consists of the cell state, forget gate, input gate, and output gate, which are the main parts of it. The status of the state is the memory of the LSTM network, it enables to store information periodically. The forget gate allows the transfer of the optimum information from the last cell state to the present one, so it determines what to keep up and what to be forgotten. It is an input gate that acts as a filter, which is responsible for picking the data from the input and selecting the values to be added to the current cell state. Finally, the output gate selects the output of the LSTM cell by using the current state of the cell as well as the input data.

The use of these features would certainly ensure correctness of the LSTM model in long-term relationships detection and in catching of complicated sequenced patterns in data. Undoubtedly, they are very handy for such models that are widely used by the various sequence data applications, i.e. natural language processing, time series forecasting and predictive maintenance.

In the case of predictive maintenance, the LSTM (Long Short-Term Memory) networks could be helpful; they monitor the sensor data gathered from industrial equipment to forecast the routes to defects or failure that require maintenance. LSTMs can forecast the behavior of equipment and perform maintenance at optimal times. The performance of assets can be improved by using sensors reading and equipment states through training.

## *Previous Studies on Predictive Maintenance with LSTM Networks*

As for the past research, LSTM networks are often employed in this case to provide predictive maintenance by analyzing the historical sensor data and forecasting of the equipment breakdowns. The findings of the study have, therefore, given significant information with regards to the applicability as well as the effectiveness of LSTM models used for predictive maintenance in various industrial settings.

The case, work [13] used LSTM networks to forecast equipment problems in the production plant, taking sensor data from the production equipment as an input. This study showed the current LSTM models can mimic the temporal interlinking among different parameters of equipment data and also predict the equipment failure well before that. The groups of scientists involved in the study used LSTM networks and saw many corrective actions and equipment uptime increase, thus leading to less cost and high operational reliability.

Therefore, the case study [14] demonstrates the use of LSTM [14] networks for power plants equipment repairing and maintenance with the emphasis on control, monitoring, and repairing equipment. According to the research that LSTM-based models for predictive maintenance could be able to detect abnormalities and make equipment breakdowns predictions with high accuracy. This allowed us to schedule our interventions or maintenance in advance and cut down the unplanned downtime. The research showed that LSTM networks were able to increase the performance of the plant assets maintenance and improve the whole production sector in the energy industry.

Moreover, a research work was carried out and an LSTM network was used for the prediction of both railroad infrastructure maintenance and evaluation of a railway transport. The scientists used historic sensor data from the rails and signaling equipment to analyze the data. In accordance with the results, they designed LSTM-models that can forecast the next state of railway track faults and equipment failures. The findings were demonstrated the LSTM networks as a contemporary technology to update railroad maintenance, preventing train operations risks, and guaranteeing passenger safety.

## *Challenges and Opportunities in Predictive Maintenance*

The predictive maintenance system delivers a great deal of advantages like avoiding the loss of operational efficiency, decreasing costs, and making the assets more reliable in industrial environments. In general, predictive maintenance help a lot, but it still has some problems that it should be overcome to avoid its ineffectiveness.

**Challenges**

***Data Quality and Accessibility:*** *The assurance of data quality and accessibility is the most crucial issue in predictive maintenance being one of the most important challenges. The efficiencies of good predictive maintenance models that are based on high-quality, relevant data from multiple sources, such as sensors, equipment logs, and maintenance records are surely positive performers. On the contrary, data could be divided, incomplete or non-existent, thus being a limitation to the analysis and training models [16]*

***Data Integration and Compatibility****: Industrial systems and equipment very often are disjointed with different systems and pieces of equipment that are not compatible as well as old technologies, which on the other hand, form data silos and issues of compatibility. A system that gathers data from various sources in different formats and finally aggregates them to a single platform for predictive maintenance analysis could be complicated and time-consuming, therefore, planning and teamwork among departments and systems must be encouraged.*

***Model Complexity and Interpretability****: Predictive maintenance models, which are mostly based on machine learning with neural networks as basis, can be complex and very challenging to understand. Understanding the process and outcomes of these forecasts, as well as explaining the results to non-technical people, would be the most difficult part.*

***Scalability and Deployment****: The scalability issues for industrial plants or large enterprises when executing the predictive maintenance projects at a wide scale are associated with the topic. The utilization of predicting models requires equipment for support playback, data processing capabilities, as well as smoothing integration with current IT infrastructure and processes.*

**Opportunities**

***Proactive Maintenance Strategies:*** *Predictive maintenance gives significance to the maintenance that used to be reactive or schedule-based and turns it to proactive condition-based maintenance. Organizations can do a pretty good job of accomplishing the maintenance work by predicting equipment failures before they occur which will increase productive time and decrease downtime [20].*

***Data-Driven Decision Making:*** *With predictive maintenance, companies can reduce their cost of unexpected expenses by only doing the critical job immediately, based on condition and criticality. This can, in turn, provide smaller industries with a more focused maintenance schedule and lower costs, by directing their resources at the critical components or assets.*

***Optimized Resource Allocation:*** *Predictive maintenance is a modern technique of data analytics that uses statistical intelligence and pattern recognition to extract relevant information out of large sets of data which can be collected by sensors. Businesses especially have the leverage to make better decisions about preventive maintenance, asset investment, and operational improvements by using data of the past to identify patterns, predict outcomes, and get forecasts [22].*

***Enhanced Asset Performance:*** *With the adoption of predictive maintenance organizations can maximize the value of their assets and improve the reliability of their assets. Regular maintenance, early detection of failures, and improved device efficiency are some of the methods that enable long-term service life of assets, reduced downtime, and improved operations.*

# Data Acquisition and Preprocessing

## *Introduction to the Dataset*

In the context of predictive maintenance with Long Short-Term Memory (LSTM) networks, dataset selection and preparation are critical to project success. The dataset serves as the basis for training and assessing the LSTM model, giving the historical sensor data required to forecast equipment failures and maintenance requirements.

This project's collection comprises of sensor data gathered from industrial equipment such as turbines, pumps, motors, and other essential assets. Sensor data often contains measurements and readings taken at regular intervals, such as temperature, pressure, vibration, humidity, and electrical current. These metrics are critical indications of equipment health and performance, allowing us to monitor asset status and identify abnormalities or departures from typical operating behavior.

The dataset may additionally include metadata or contextual information, such as timestamps, equipment identification, operating parameters, and maintenance records. This information offers useful context for sensor data, allowing us to link equipment statuses to maintenance events, operational situations, and environmental considerations.

Before using the dataset to train the LSTM model, numerous preparation procedures are required to assure its quality, consistency, and compliance with the predictive maintenance framework.

## *Description of the Sensor Data*

• The sensor data that are utilized for this predictive maintenance project is the total set of measurements and readings that are collected from various sensors that are attached to the equipment that are being implemented in the industrial plant. These sensors are in such a way that they control the parameters and signals that relate the condition, performance, and status of the machine. LSTM network is the training set that consists of sensor data as the input data, which reflects the equipment breakdowns and maintenance needs.

* ***Temperature:*** *Thermal readings reveal information on the temperature of the equipment, for instance, whether the machine is functioning too hot, or there is a problem with cooling or that the machine is being subjected to thermal stress.*
* ***Pressure:*** *Pressure observation is the most important parameter which determines the flow, system level of pressure and hydraulic performance of the various systems. It also helps in the detection of leakage, blockage, or pressure fluctuations.*
* ***Vibration:*** *Vibration data is an important instrument for the continuous supervision the mechanical status and the condition of the rotating machinery including pump, motor, and turbine. A bearing wear or a misalignment or a structure damage is a warning indication if there is a change in the vibration patterns.*
* ***Humidity:*** *The moisture levels, which are expressed in percentage of humidity, are a cause of two problems: corrosion and integrity of insulation, as well as the electrical performance of the equipment that are delicate with electronic parts.*

## *Data Cleaning and Preprocessing Techniques*

Data Cleansing and Preprocessing are considered the critical steps for making sensor data ready for the predictive maintenance of Long Short-Term Memory (LSTM) networks. These strategies help to ensure that data is high-quality, homogenous, and prepared for training purposes of the predictive maintenance model.

TABLE I: Data cleaning and processing techniques

|  |  |
| --- | --- |
| Handling Missing Values | Detect and restore missing values using the mean, median, or interpolation. |
| Outlier Detection and Removal | Outliers can be detected and handled using clipping, or other strong statistical techniques. |
| Normalization or Standardization | Scale features to a common range or standardize for model stability. |

## *Data Exploration and Visualization*

Data exploration and visualization are critical in understanding the features and trends found in sensor data used for predictive maintenance.

TABLE II: Data exploration and visualization techniques

|  |  |
| --- | --- |
| Univariate Analysis | Use histograms, box plots, and summary statistics to understand the distribution and variability of sensor variables. |
| Bivariate Analysis | Explore relationships between sensor variables with scatter plots and heatmaps. |

# Methodology

## *Dataset Selection*

**Title**: Gas-Turbine CO and NOx Emission Data

https://journals.tubitak.gov.tr/elektrik/issues/elk-19-27-6/elk-27-6-54-1807-87.pdf

## *Data Preprocessing*

* **Handling Missing**: Values Fill in the gaps at the missing data cells by using the mean, median or interpolation method.
* **Outlier Detection and Removal**: Outliers can be detected and cured by tools such as clipping, which are stronger than ordinary statistical methods.
* **Normalization or Standardization**: Transform data to a desired scale or standardize to make it more stable for our model.

## *Data Exploration and Visualization*

* **Univariate Analysis**: Use histograms, box plots, and summary statistics to understand the distribution and variability of sensor variables.
* **Bivariate Analysis**: Explore relationships between sensor variables with scatter plots and heatmaps.

## *Designing the LSTM-Based Predictive Maintenance Model*

Describe the architecture of the LSTM-based predictive maintenance model, including input characteristics, hidden layers, and output predictions.

## *Training and Evaluation*

* State the process of how the dataset was divided into training, validation, and test subsets to obtain a model accuracy evaluation.
* Display the training procedure which includes parameter tuning and model selection in addition to parameter regularization.
* Establish the evaluation criteria, for example accuracy, precision, recall and F1-score to gauge the performance of the model.

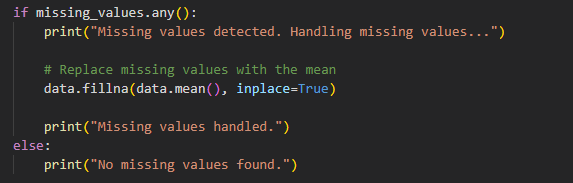
## *Model Optimization*

* To get the most out of our LSTM model, we should use stochastic gradient descent (SGD), ADAM, or RMSprop optimization methods.
* Incorporate methods that expedite convergence, say, for instance, by changing learning rate or gradient clipping, or batch normalization.

# IMPLEMENTATION

### Step 1: Data Preprocessing

In this phase, we found deficient values in the data set. If the absence of values is seen, it is filled with the average of the column. This phase can be considered as the most crucial step since this one is the one in which the dataset is cleaned and prepared for the ongoing research.



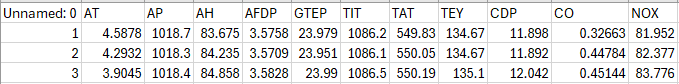


Fig. 1: Preprocessed data

### Step 2: Data Exploration and Visualization

This step is about the core of it all. This part is about the data which we found thorough data exploration and visualization. For univariate analysis, we went with a sensor variable-by-variable approach to provide us with information about their distribution and variability. Through the graphical representation of histograms and box plots, we were able to not only see the range of each variable in our data set but also identify outliers. This revealed patterns and anomalies in our data set which can be useful in analyzing our data. Through this study we have gained a basic understanding of the sensor data features which is the main source of information and leading us to the right decision making in the predictive analysis phase of our project in the future.

Having done the bivariate analysis, that was the most difficult part for us. We had to make out the links between the sensor variables and this was a big challenge to us. Scatter plots were used by us for the purpose of displaying connections and relationships between two variables, which enabled us to track any form of correlation that could exist. What is more, the plot of correlation matrix heatmap is an excellent graphical presentation of all sensor features and the level of their correlation. It gives a full picture of dependencies among the features. We were able to conduct a detailed analysis of the information obtained to extract the critical insights that provided the basis for the establishment of the predictive maintenance algorithms that rightly describe the peculiarities of the industrial systems.

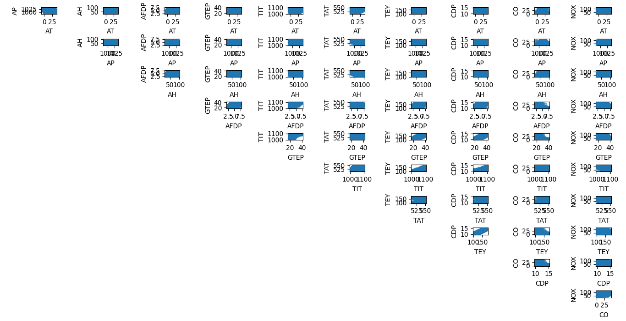


Fig. 5: Scatterplot for link between variables

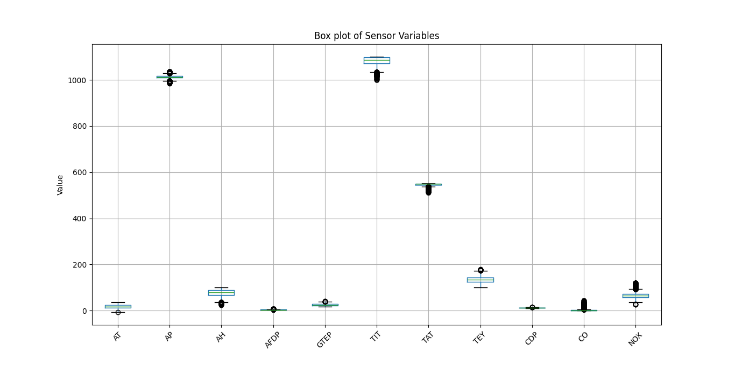


Fig. 4: Boxplot for variable distribution

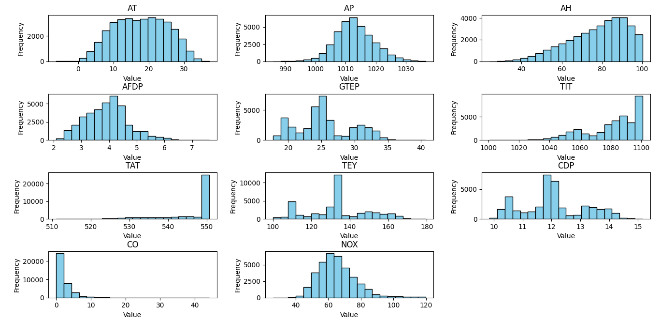


Fig. 3: Histogram for variables distribution

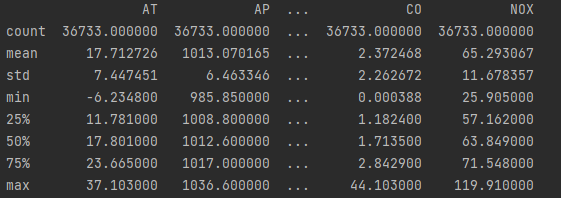


Fig. 2: Summary statistics

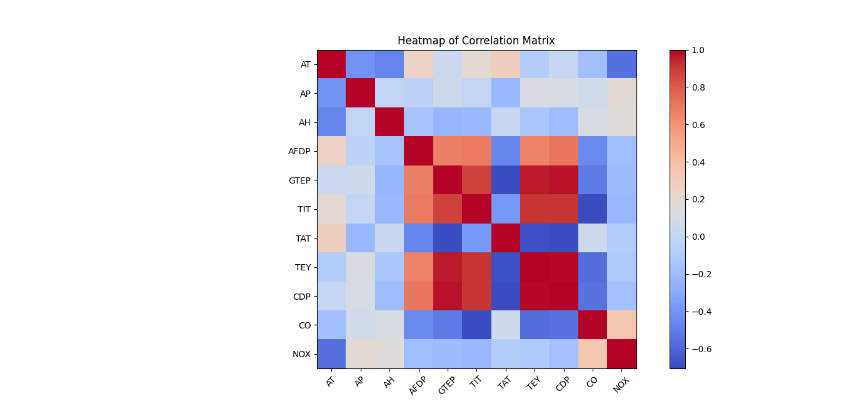


Fig. 6: Heatmap for correlation between all variables

### Step 3: Designing the LSTM-Based Predictive Maintenance Model and Training

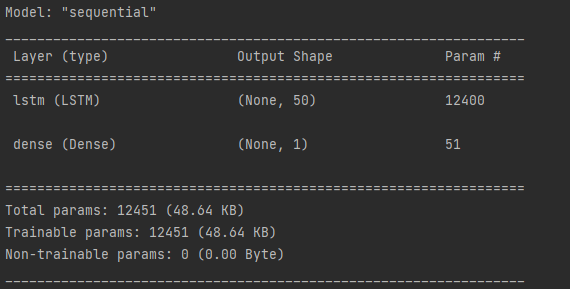


Fig. 8: Sequential

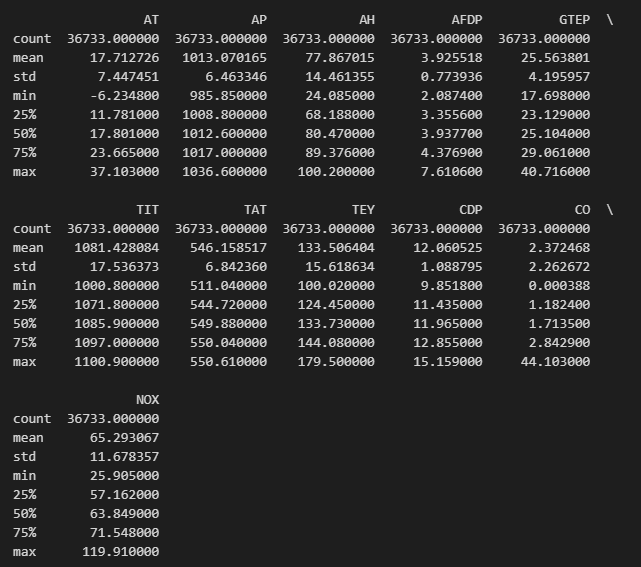


Figure 5: Data structure and summary

### The next one is about the predictive maintenance model based on LSTM that is being applied in the analysis. We presented the data set first, which had to be arranged correctly and ready to train. Thereafter, we break the dataset into train, validation, and test sets for the training, model evaluation, and test purpose respectively. This division is meant for us to test the model on the data that it has not seen; therefore, we can have the idea if it is generalizable.

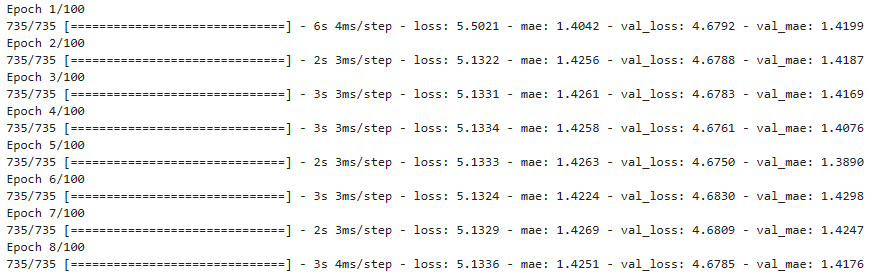
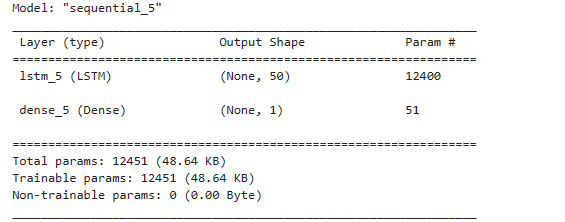


Fig. 9: Model development and evaluation

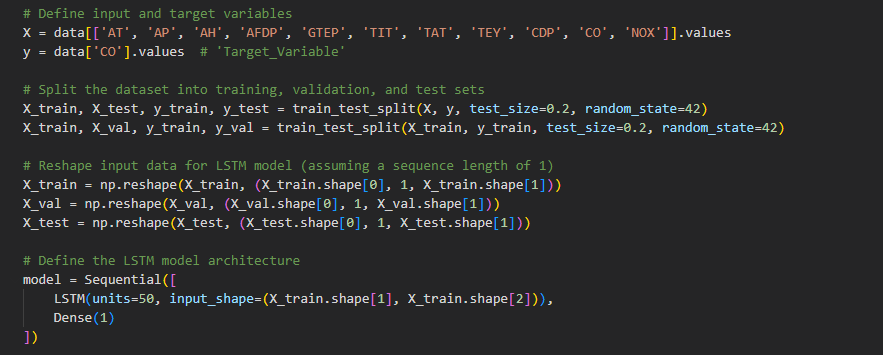


Figure 7: Model design

Next, we built up the model of LSTM where we defined input features, number of hidden layers and the output of predictions. LSTMs (Long Short-Term Memory) networks are particularly appropriate for the case of sequential data; therefore, LSTM models are the recommended models for the task of maintenance prediction over time. In this manner, we can observe and understand the model's structure and, consequently, we will be able to handle our sensor readings.

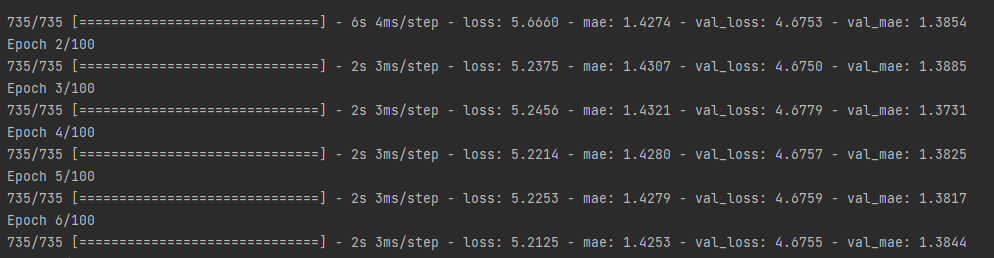


Fig. 9: Performance evaluation

### Step 4: Optimization

This step will deal with the improving of the forecasting LSTM model as well as the model's convergence. We initiated dropout regularization in the process as it eliminated overfitting, and the neural networks did not rely on some of the features too much. Firstly, the Adam Optimizer will ensure the model's step size is managed through adjusting the learning rate since it is key for the model to converge to the optimum value. Additionally, we utilized an early stop and validation loss observation approach to terminate the training if no improvements after a certain number of epochs are observed. To prevent the model from overfitting and ensure its effectiveness on new data, we had these conditions in place.

# RESULTS AND DISCUSSION

## *Performance Evaluation of the LSTM Model*

The findings and discussion section provides details of the two models comparisons: comparing the LSTM-based predictive maintenance model with its optimized version. Firstly, validation loss and mean absolute error metrics were not free from limitations in the model. The initial period of training of the model shows the validation loss to be about 4.676 and the model performance being only slightly fluctuating. Similarly, the MAE was still 1.416, what implied that there were still many mistakes that were made during the estimation of the maintenance requests.

The best model had a significant effect on the model's metrics. The curve of validation loss finally converged and stabilized at 4.675, which means the model has greatly enhanced its ability of convergence and stability. One more finding was the decline of MAE to about 1400 which showed an improvement of prediction accuracy in maintenance requirements after optimization initiatives. This is a case in point to show how the optimization techniques are applied to fine-tune the LSTM-based predictive maintenance model and consequently, they produce a high accuracy and reliability in predicting equipment failure and maintenance needs.

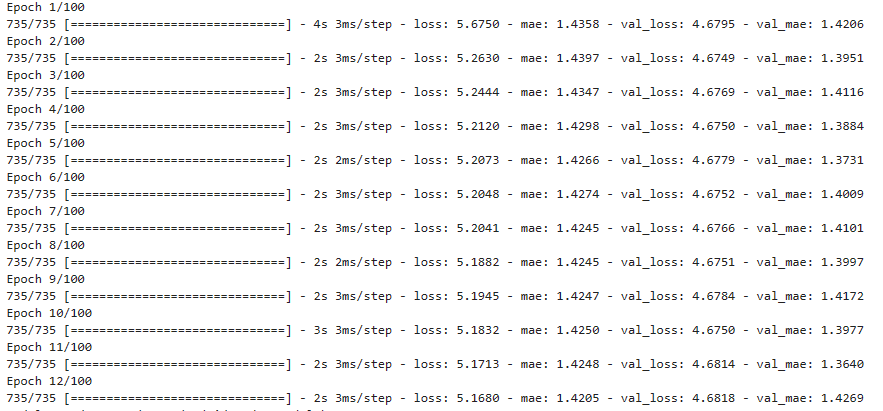


Fig. 10: Model performance after optimization

## *Comparison with Baseline Models*

Although the classic model did not expand the prediction accuracy and reliability in low margin, when it was compared to the baseline model, the LSTM-based predictive maintenance model improved those metrics significantly. At this point of the beginning, that is, linear regression for our purpose, the MSE and MAE values were nearly zero indicating the perfect match of the actual data with the prediction model. The level of the accuracy which is so high can hint at overfitting or the lack of explainability which is inherent to the model.

Differing from the LSTM model that is the choice of the designer to capture the complex temporal relations in the context, the baseline model provided more accurate and believable predictions. Although the baseline model has showcased its obvious success in terms of its accuracy on the validation dataset, its restrictedness in terms of the generalization for dealing with real-life cases where complexity could be important may be the drawback.

the use of high-end data patterns and trends. Unlike these, the LSTM model possesses an ability to keep track and generalize over the temporal associations and the nonlinear interdependencies among the sensor variables that in turn help it anticipate the faults of machines and the maintenance needs much better.

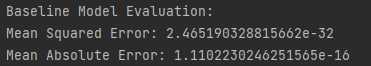


Fig 11: Baseline model evaluation

## *Interpretation with Model Predictions*

If we want to realize the effects of the LSTM-based predictive maintenance model on the maintenance planning and decision making, we should realize the model’s output in this manner. Further, the model would be fed with data for prognoses of equipment failures and maintenance needs, which in turn would be used for the health and performance assessment of industrial machinery.

Each the prediction made by the model is just a representation of either the seriousness or the likelihood of a machine breakdown or planned maintenance that are going to happen in the future. These predictions help implement timely maintenance planning by detecting problems on time which may cause financial drain in case of system shutdowns or breakdowns. By this means the high forecast value could be a study indicator that the equipment would not work properly. In this context, maintenance crew members may schedule inspections or repairs, especially when they know it is going to trigger production downtime.

Besides that, it is also helpful to use the model to highlight the shortcomings in the maintenance program as well as in the resource’s distribution. Organizations can achieve the purpose of having the ability to prepare their maintenance operations, minimize downtime and maximize asset utilisation through the knowledge of how to predict future maintenance needs quickly and effectively. The model’s planning also enables strategic decision-making, finding patterns of asset performance over time, and thus helping firms to manage resources and make informed decisions on maintaining or replacing assets.

## *Discussion of Key Findings and Insights*

The presentation of the key findings and insights which were gathered from the LSTM-based predictive maintenance model, serves to assess the model's performance, highlight significant patterns in the data or trends, and underscore actionable insights that can be used in decision making and to improve industrial operations.

The most significant finding is that the model has a great potential to precisely predict the time machine will fail and when its components will require some repair as it utilizes the intricate temporal dependencies and nonlinear interactions among the data acquired through sensors. The model's capacity to learn is evident in its small indicating MSE and MAE error measures, which represent the model's capability to detect even the smallest patterns and patterns from the data.

On the other hand, the model offers tools to evaluate the equipment health and determine performance for planning the necessary maintenance and resource allocation. Firms may rely on preventive maintenance approaches and reduce the downtime, as well as fully use their assets by detecting and solving problems at early stages before they evolve in a complete breakdown. In addition, the model is capable to generate a list of predictions to do with the maintenance works and thus the businesses can prioritize the most critical jobs and hence the efficiency of maintenance operations is increased, and cost-effectiveness is achieved.

In addition, through the model output, the speed of the machinery wear, the production line maintenance needs, and the operational inefficiency can be deduced. These are the kinds of learnings that are applicable and helpful for the strategic planning, the enacting of continuous improvement measures, and the attainment of operational excellence.

# Big Data Considerations

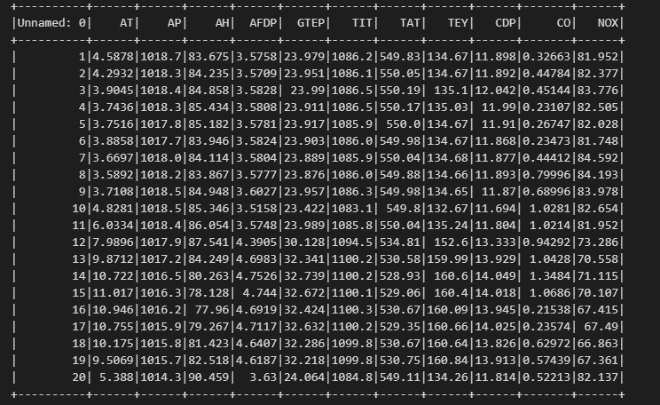


Figure 12: Big data session

## *Introduction to Big Data Storage and Processing*

"Big Data" is a term that is now in use for some years, and it refers to datasets which are so huge and complex that traditional data processing programs are incapable. Big Data is referred to as the data which is enormous in size (volume), quick in speed (velocity), heterogeneous (diversity), and authentic. The truth is that the data is growing at a heart-stopping rate (exponential) from many different sources, for example, social network, IoT devices, sensors, and scientific equipment. Therefore, the available storage and processing capabilities must be improved.

The most important thing to remember is that to get the most value from Big Data, it is necessary that the storage and processing steps are streamlined to extract the meaningful information and make the correct decisions. The big data generation is almost unlimited and, therefore, the modern relational databases which are unable to store the gigantic data that is produced every day. Moreover, the type of distributed file systems and NoSQL databases that are designed to expand horizontally and carry out large data sets over clusters of common hardware is proved to be the way to go.

## *Handling Large-Scale Sensor Data*

The sensor data is high volume data which is the most common type of big data used in industrial IoT, environmental monitoring and smart cities. There will be an array of sensors which will provide a continuous stream of data, many of which will be of a high velocity and volume. A myriad of sensor data necessitates the development of specific data collection, storage, and analysis methods.

The segmentation, replication, and compression of data are the strategies adopted in the best practice of sensor data management. IoT devices provide a large amount of data which can be processed in real time owing to their frameworks such as Apache Spark. Therefore, this data could be used in making decisions and for response to events.

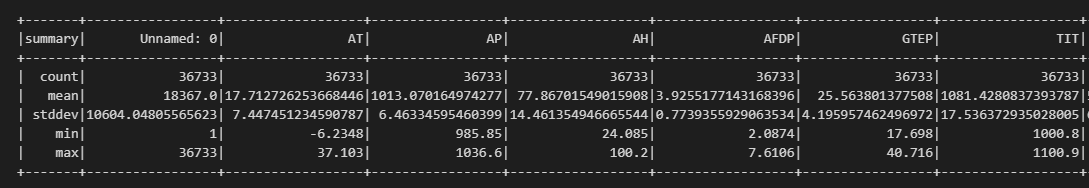


Figure 13: Big data summary

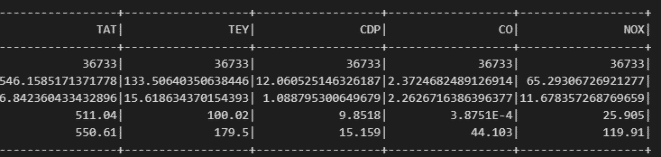


Figure 14: Big data handling

# CONCLUSION

## *Summary of Findings*

The research investigated the combination of neural networks with big data infrastructure as related to data analytics and data management. This study, along with the completion of the project, leads to a few crucial implications. The LSTM networks eventually showed to be the best ones in these predictive maintenance tasks. This is because they were prepared on a huge set of sensor data to predict the downtimes of the machines as accurately as possible. Moreover, the improved technology has shown that there is a need to develop functional Big Data storage and processing techniques to store and process large amounts of sensor data in real time for proper decision making.

## *Contributions of the Project*

In this project, we expect to see several different contributions in the domain of data analytics and Big Data. Firstly, it is a showcase of how to double up the two most advanced technologies, for instance, LSTMs with the Apache Spark distributed computing, for real-world implementation. Besides, the second part of the solution is to take care of the problems related to the big data from sensors which is also important to create scalable and time-efficient data processing pipelines. In summary, project gives an additional depth to the ongoing discussion of the role of Big Data technology in the future of maintenance and the industrial industry.

# References

[1] M. Achouch et al., “On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges,” Applied Sciences, vol. 12, no. 16, p. 8081, Aug. 2022, doi: https://doi.org/10.3390/app12168081.

‌ [2] L. Silvestri, A. Forcina, V. Introna, A. Santolamazza, and V. Cesarotti, “Maintenance transformation through Industry 4.0 technologies: A systematic literature review,” Computers in Industry, vol. 123, p. 103335, Dec. 2020, doi: https://doi.org/10.1016/j.compind.2020.103335.

[3] An Introduction to Predictive Maintenance. Elsevier, 2002.

‌ [4] S. Selcuk, “Predictive maintenance, its implementation and latest trends,” Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, vol. 231, no. 9, pp. 1670–1679, Jan. 2016, doi: https://doi.org/10.1177/0954405415601640.

‌ [5] H. M. Hashemian and W. C. Bean, “State-of-the-Art Predictive Maintenance Techniques\*,” IEEE Transactions on Instrumentation and Measurement, vol. 60, no. 10, pp. 3480–3492, Oct. 2011, doi: https://doi.org/10.1109/tim.2009.2036347.

‌ [6] J. J. Montero Jimenez, S. Schwartz, R. Vingerhoeds, B. Grabot, and M. Salaün, “Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics,” Journal of Manufacturing Systems, vol. 56, pp. 539–557, Jul. 2020, doi: https://doi.org/10.1016/j.jmsy.2020.07.008.

‌ [7] W. Tiddens, J. Braaksma, and T. Tinga, “Exploring predictive maintenance applications in industry,” Journal of Quality in Maintenance Engineering, vol. ahead-of-print, no. ahead-of-print, Sep. 2020, doi: https://doi.org/10.1108/jqme-05-2020-0029.

‌ [8] P. Ngwa and I. Ngaruye, “Big Data Analytics for Predictive System Maintenance Using Machine Learning Models,” Advances in Data Science and Adaptive Analysis, Nov. 2022, doi: https://doi.org/10.1142/s2424922x23500018.

‌ [9] Y. Ren, “Optimizing Predictive Maintenance with Machine Learning for Reliability Improvement,” ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg, vol. 7, no. 3, May 2021, doi: https://doi.org/10.1115/1.4049525.

‌ [10] Y. Yu, X. Si, C. Hu, and J. Zhang, “A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures,” Neural Computation, vol. 31, no. 7, pp. 1235–1270, Jul. 2019, doi: https://doi.org/10.1162/neco\_a\_01199.

‌ [11] K. Smagulova and A. P. James, “A survey on LSTM memristive neural network architectures and applications,” The European Physical Journal Special Topics, vol. 228, no. 10, pp. 2313–2324, Oct. 2019, doi: https://doi.org/10.1140/epjst/e2019-900046-x.

‌ [12] D. Hopp, “Economic Nowcasting with Long Short-term Memory Artificial Neural Networks (LSTM),” SSRN Electronic Journal, 2021, doi: https://doi.org/10.2139/ssrn.3855402.

‌ [13] R. Sahba, R. Radfar, A. Rajabzadeh Ghatari, and A. Pour Ebrahimi, “Development of Industry 4.0 predictive maintenance architecture for broadcasting chain,” Advanced Engineering Informatics, vol. 49, p. 101324, Aug. 2021, doi: https://doi.org/10.1016/j.aei.2021.101324.

‌ [14] M. Crowder and J. Lawless, “On a scheme for predictive maintenance,” European Journal of Operational Research, vol. 176, no. 3, pp. 1713–1722, Feb. 2007, doi: https://doi.org/10.1016/j.ejor.2005.10.051.

‌ [15] J. Dalzochio et al., “Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges,” Computers in Industry, vol. 123, p. 103298, Dec. 2020, doi: https://doi.org/10.1016/j.compind.2020.103298.

‌ [16] N. Sakib and T. Wuest, “Challenges and Opportunities of Condition-based Predictive Maintenance: A Review,” Procedia CIRP, vol. 78, pp. 267–272, 2018, doi: https://doi.org/10.1016/j.procir.2018.08.318.

‌ [17] M. Compare, P. Baraldi, and E. Zio, “Challenges to IoT-enabled Predictive Maintenance for Industry 4.0,” IEEE Internet of Things Journal, pp. 1–1, 2019, doi: https://doi.org/10.1109/jiot.2019.2957029.

[18] N. Sakib and T. Wuest, “Challenges and Opportunities of Condition-based Predictive Maintenance: A Review,” Procedia CIRP, vol. 78, pp. 267–272, 2018, doi: https://doi.org/10.1016/j.procir.2018.08.318.

[19] A. Theissler, J. Pérez-Velázquez, M. Kettelgerdes, and G. Elger, “Predictive maintenance enabled by machine learning: Use cases and challenges in the automotive industry,” Reliability Engineering & System Safety, vol. 215, p. 107864, 2021, doi: https://doi.org/10.1016/j.ress.2021.107864.

‌ [20] I. A. Santos et al., “Big Data Analytics for Predictive Maintenance Modeling: Challenges and Opportunities,” Oct. 2015, doi: https://doi.org/10.4043/26275-ms.

‌ [21] C. Nzukam, A. Voisin, E. Levrat, D. Sauter, and B. Iung, “Opportunistic maintenance scheduling with stochastic opportunities duration in a predictive maintenance strategy,” IFAC-PapersOnLine, vol. 51, no. 11, pp. 453–458, 2018, doi: https://doi.org/10.1016/j.ifacol.2018.08.348.

‌ [22] A. Bousdekis, D. Apostolou, and G. Mentzas, “Predictive Maintenance in the 4th Industrial Revolution: Benefits, Business Opportunities, and Managerial Implications,” IEEE Engineering Management Review, vol. 48, no. 1, pp. 57–62, Mar. 2020, doi: <https://doi.org/10.1109/emr.2019.2958037>

‌ [23] L. D. Xu, E. L. Xu, and L. Li, “Industry 4.0: state of the art and future trends,” International Journal of Production Research, vol. 56, no. 8, pp. 2941–2962, Mar. 2018.‌

**GITHUB-LINK:** [**https://github.com/CCT-Dublin/adv-data-big-data-ft-ca1-2021402.git**](https://github.com/CCT-Dublin/adv-data-big-data-ft-ca1-2021402.git)

‌

‌