**OPTIMIZATION OF MACHINE DOWNTIME**

**DONE FOR**

****

Project Report Submitted in partial fulfilment of the requirement of

**PONDICHERRY UNIVERSITY** for the award of the degree of

**MASTER OF BUSINESS ADMINISTRATION(DATA ANALYTICS)**

By

**JOICY**

(Reg. No. 22401016)

Under the Guidance of

**DR. A. KARTHIGEYAN**

Assistant Professor,

Department of Management Studies

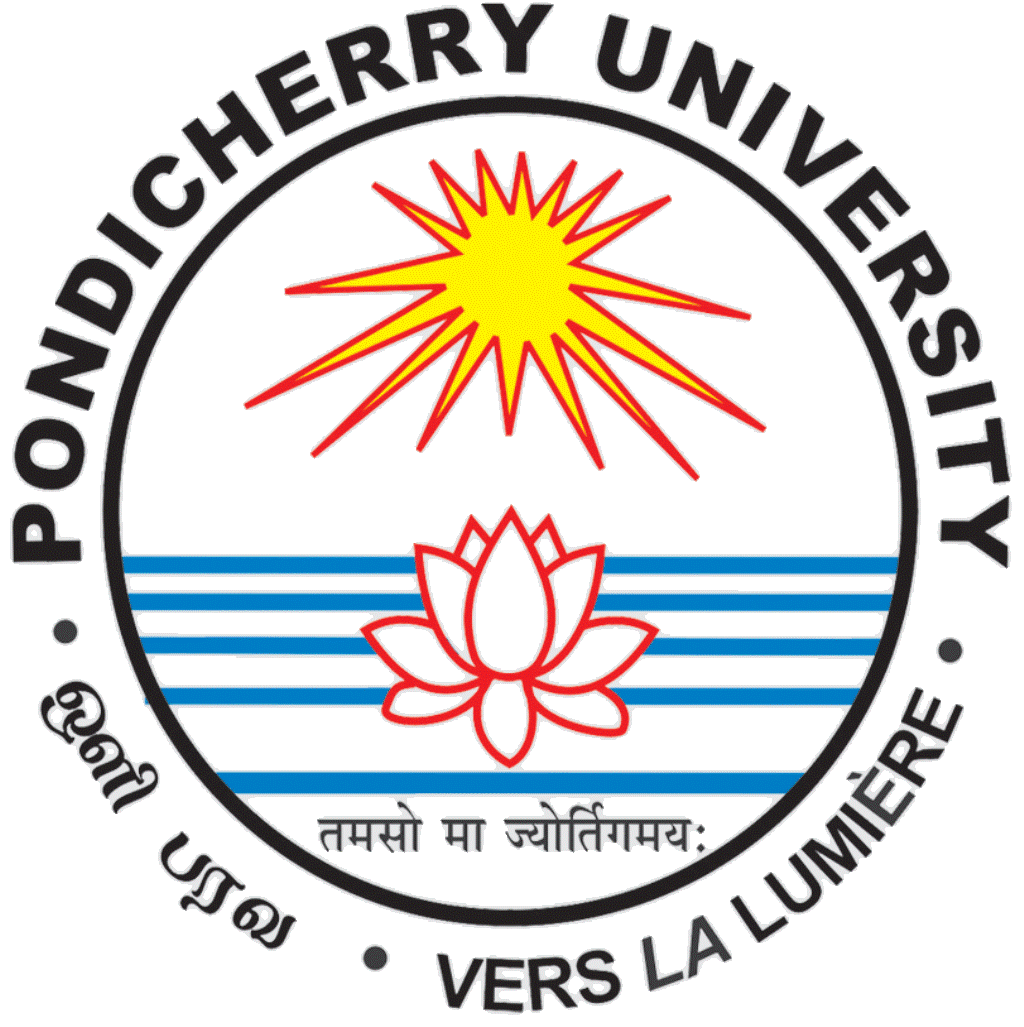
Pondicherry University

&

**MR. SHARAT MANIKONDA**

VP, Analytics,

360Digi TMG

****

**DEPARTMENT OF MANAGEMENT STUDIES**

**SCHOOL OF MANAGEMENT**

**PONDICHERRY UNIVERSITY**

**PONDICHERRY-605014**

**JANUARY – FEBRUARY 2023**



**DECLARATION**

I hereby declare that the project titled, **“OPTIMIZATION OF MACHINE DOWNTIME”**is an original work done by me under the guidance of **DR. A. KARTHIGEYAN, Assistant Professor,** Department of Management Studies, Pondicherry University and **MR. SHARAT MANIKONDA**, VP, Analytics, 360Digi TMG. This project or any part thereof has not been submitted for any Degree / Diploma / Associateship / Fellowship / any other similar title or recognition to this University or any other University.

I take full responsibility for the originality of this report. I am aware that I may have to forfeit the degree if plagiarism has been detected after the award of the degree. Notwithstanding the supervision provided to me by the Faculty Guide, I warrant that any alleged act(s) of plagiarism in this project report are entirely my responsibility. Pondicherry University and/or its employees shall under no circumstances whatsoever be under any liability of any kind in respect of the aforesaid act(s) of plagiarism.

**Joicy**

Reg. No. 22401016

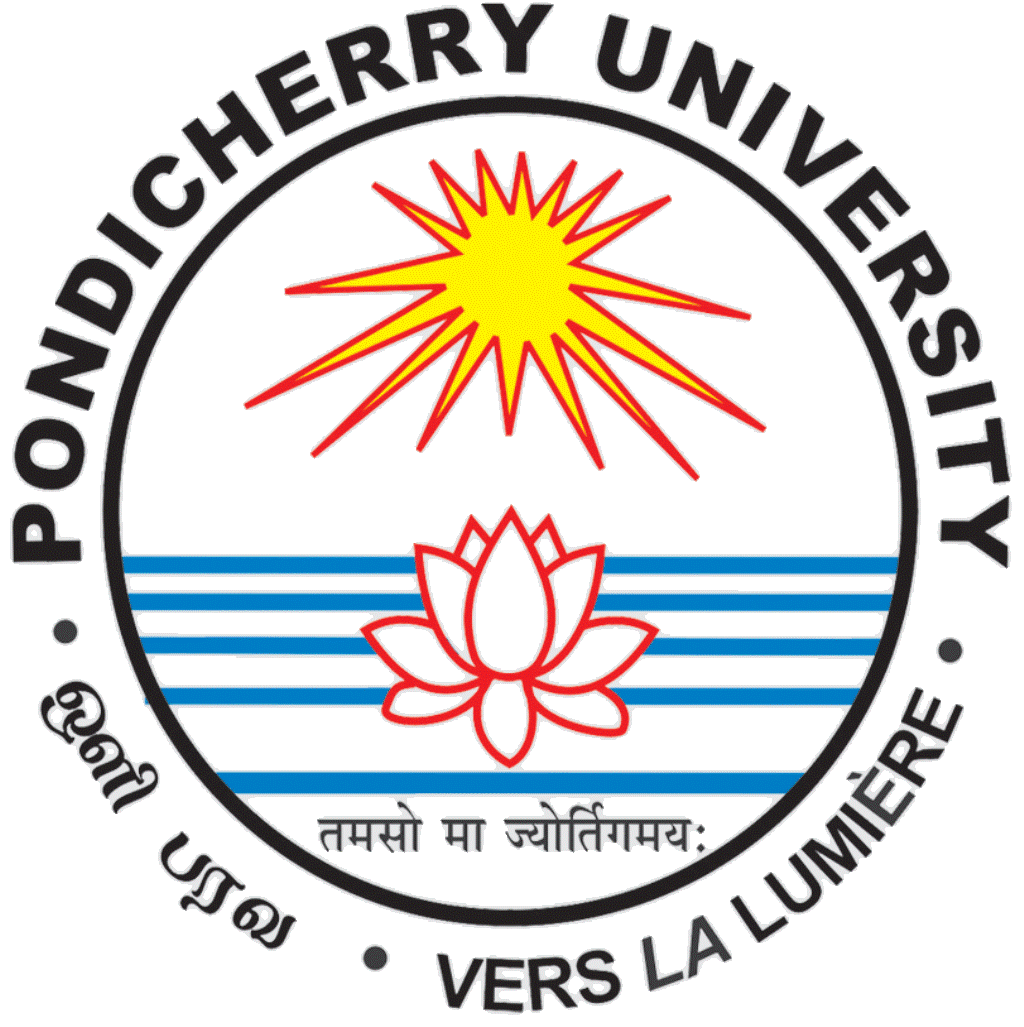
II MBA   
Department of

Management Studies

Pondicherry University

Date:

Place: Pondicherry

****

**DEPARTMENT OF MANAGEMENT STUDIES**

**SCHOOL OF MANAGEMENT**

**PONDICHERRY UNIVERSITY**

**PONDICHERRY-605014**

**CERTIFICATE**

This is to certify that this project report entitled **“Optimization of Machine Downtime”** done for 360 DigiTMT is submitted by **Joicy (Reg.No:22401016),**  II MBA to the **DEPARTMENT OF MANAGEMENT STUDIES, SCHOOL OF MANAGEMENT, PONDICHERRY UNIVERSITY** in partial fulfilment of the requirements for the award of the degree of **MASTER OF BUSINESS ADMINISTRATION(DATA ANALYTICS)** and is a record of an original and bonafide work done under the guidance of **DR A. KARTHIGEYAN**, Assistant Professor, Department of Management Studies, Pondicherry University. This report has not formed the basis for the award of any degree, diploma, associateship, fellowship or other similar title to the candidate and that the report represents an independent and original work on the part of the candidate.

**DR. A. KARTHIGEYAN**

Assistant Professor

and Project Guide

Department of Management Studies

**Dr. R. Kasilingam**

Professor and Head

Department of Management Studies

Date:

Place: Pondicherry, 605 014

**ACKNOWLEDGMENTS**

           I express my deep sense of gratitude to my guide Dr. A. Karthigeyan, Professor, Department of Management Studies, Pondicherry University and Mr. Sharat Manikonda, VP, Analytics, 360Digi TMG, for the valuable guidance, constant encouragement and creative suggestions offered during the course of the project and also in the preparation of this report. I extend my sincere thanks to Dr. R. Kasilingam, Professor and Head, Department of Management Studies, Pondicherry University, for providing guidance and facilities for the project.

I also extend my sincere thanks to all other faculty members of the Department of Management Studies, Pondicherry University, my parents and also my friends for their constant support. Above all, I thank the God Almighty for giving the grace without which it would not have been possible to complete this project on time.

**Joicy**

Reg. No. 22401016

II MBA   
Department of Management Studies

Pondicherry University

**TABLE OF CONTENTS**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| I. | CERTIFICATE | | | | | | |  |
| II. | CERTIFICATE FROM ORGANIZATION | | | | | | |  |
| III. | DECLARATION | | | | | | |  |
| IV. | CERTIFICATE BY GUIDE AND HOD | | | | | | |  |
| V. | ACKNOWLEDGEMENTS | | | | | | |  |
| VI. | TABLE OF CONTENTS | | | | | | |  |
| VII. | LIST OF FIGURES | | | | | | |  |
| VIII. | CHAPTER | | | | | | |  |
|  |  | | | | | | |  |
|  | 1. | INTRODUCTION | | | | | | 08-13 |
|  |  | 1.1 | INTRODUCTION | | | | | 08 |
|  |  | 1.2 | STATEMENT OF PROBLEM | | | | | 09 |
|  |  | 1.3 | OBJECTIVES. | | | | | 09-10 |
|  |  | 1.4 | METHODOLOGY. | | | | | 10-12 |
|  |  | 1.5 | LIMITATIONS OF THE STUDY. | | | | | 13 |
|  |  |  |  |  |  |  |  |  |
|  | 2. | PROFILE OF INDUSTRY AND COMPANY | | | | | | 14-15 |
|  |  | 2.1 | ABOUT THE INSTITUTE | | | | | 14 |
|  |  | 2.2 | MISSION. | | | | | 14 |
|  |  | 2.4 | OBJECTIVES. | | | | | 16- |
|  |  |  |  |  |  |  |  |  |
|  | 3. | CONCEPTUAL FRAMEWORK ON TOPIC | | | | | | 16-20 |
|  |  | 3.1 | PREDICTIVE MAINTENANCE | | | | | 16-17 |
|  |  | 3.2 | MACHINE LEARNING IN PREDICTIVE | | | | |  |
|  |  |  | MAINTENANCE | | | | | 17-19 |
|  |  | 3.3 | KEY CONCEPTS AND TERMINOLOGY | | | | | 19-20 |
|  |  |  |  |  |  |  |  |  |
|  | 4. | DATA ANALYSIS AND MODELLING | | | | | | 21- |
|  |  | 4.1 | DATA PREPROCESSING | | | | | 21-22 |
|  |  | 4.2 | EXPLORATORY DATA ANALYSIS(EDA) | | | | | 22-26 |
|  |  | 4.3 | MACHINE LEARNING MODELLING | | | | | 26-34 |
|  |  |  |  |  |  |  |  |  |
|  | 5. | FINDINGS AND CONCLUSION. | | | | | | 35-37 |
|  |  | 5.1 | FINDINGS. | | | | | 35-36 |
|  |  | 5.2 | CONCLUSION. | | | | | 36-37 |
|  |  |  |  |  |  |  |  |  |
| IX. | BIBLIOGRAPHY | | | | | | |  |
|  |  | | | | | | |  |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIG. NO.** | **NAME** | **PAGE NO.** |
| 1 | Data Preprocessing\_1 | 21 |
| 2 | Data Preprocessing\_2 | 21 |
|  | Checking for Duplicates | 2 |
| 4 | Encoding data | 22 |
| 5 | Overall Analysis | 22 |
| 6 | Unique Assembly Lines | 23 |
| 7 | Downtime counts and Percentages for each assembly line | 23 |
| 8 | Number of Unique Machine IDs | 23 |
| 9 | Downtime Counts and Percentages for each Machine ID | 24 |
| 10 | Code snip for correlation analysis | 24 |
| 11 | Correlation Matrix | 25 |
| 12 | Correlation Heatgraph | 25 |
| 13 | Encoding variables and Data Splitting | 26 |
| 14 | Algorithm Selection | 27 |
| 15 | Model Evaluation | 28 |
| 16 | Hyperparameter Tuning | 29 |
| 17 | Evaluating Model with Best Parameters | 31 |
| 18 | Feature Importance Analysis | 32 |

**CHAPTER 1: INTRODUCTION**

**1.1 INTRODUCTION**

In industrial settings, machine downtime refers to the period during which a machine or equipment is non-operational due to various reasons such as mechanical failures, breakdowns, maintenance activities, or unplanned stoppages. Machine downtime poses significant challenges for manufacturers, disrupting production schedules, increasing maintenance costs, and impacting overall productivity and profitability.

Optimizing machine downtime has emerged as a critical objective for industrial organizations seeking to enhance operational efficiency, reduce costs, and improve competitiveness. By minimizing downtime and maximizing equipment availability, manufacturers can ensure smooth production processes, meet customer demands, and maintain a competitive edge in the market.

Predictive maintenance, an advanced maintenance strategy enabled by machine learning and data analytics, has gained traction as a proactive approach to optimize machine downtime. Predictive maintenance involves the use of historical data, sensor readings, and predictive models to anticipate equipment failures before they occur. By analyzing patterns and trends in machine performance data, predictive maintenance models can identify early indicators of potential failures, enabling timely interventions to prevent or mitigate downtime.

The need for predictive maintenance stems from the limitations of traditional maintenance practices, such as reactive and preventive maintenance. Reactive maintenance, which involves repairing machines only after they have failed, is often costly and disruptive, leading to prolonged downtime and increased repair costs. On the other hand, preventive maintenance, which relies on scheduled inspections and maintenance tasks, may not be effective in preventing unexpected failures or optimizing maintenance schedules.

By leveraging predictive maintenance models, manufacturers can transition from reactive and preventive maintenance strategies to a proactive approach that focuses on optimizing maintenance schedules, reducing downtime, and improving equipment reliability. By anticipating and addressing potential failures before they occur, predictive maintenance enables organizations to minimize disruptions, enhance operational efficiency, and achieve greater competitiveness in the market.

**1.2 STATEMENT OF PROBLEM**

Unplanned machine downtime is occurring in the manufacturing process of vehicle fuel pumps. This downtime leads to a loss of productivity and can disrupt production schedules. The occurrence of unexpected downtime poses significant challenges for the manufacturing facility, resulting in decreased operational efficiency, increased maintenance costs, and potential revenue losses. Addressing this problem is crucial for maintaining smooth production operations, meeting customer demands, and ensuring the profitability of the manufacturing process.

The need to minimize machine downtime and its associated impacts on productivity is evident. Traditional maintenance approaches, such as reactive and preventive maintenance, may not be sufficient to prevent unexpected failures and optimize maintenance schedules effectively. There is a pressing need to implement proactive maintenance strategies that can anticipate potential machine failures before they occur, enabling timely interventions to prevent or mitigate downtime.

In light of these challenges, the primary objective of this project is to develop a predictive maintenance model for vehicle fuel pump manufacturing processes. By leveraging historical data on machine performance and failure incidents, our aim is to build a predictive model that can anticipate potential failures and optimize maintenance schedules to minimize downtime and maximize productivity. This predictive maintenance model will enable proactive maintenance interventions, thereby improving operational efficiency, reducing maintenance costs, and enhancing overall equipment reliability in the manufacturing facility.

**1.3 OBJECTIVES**

* Identify key factors contributing to unplanned machine downtime in the manufacturing process and analyze their impact on production schedules and productivity.
* Develop a predictive maintenance model for the manufacturing process of vehicle fuel pumps to anticipate potential machine failures before they occur.
* Optimize maintenance schedules based on predictive insights from the developed model to minimize unplanned machine downtime and maximize equipment availability.
* Evaluate the effectiveness of the predictive maintenance model in minimizing unplanned machine downtime and improving operational efficiency in the manufacturing facility.
* Provide recommendations for continuous improvement and refinement of maintenance strategies based on insights gained from the predictive maintenance model and performance evaluation results.

**1.4 METHODOLOGY**

**Data Collection and Understanding:**

Primary data was received directly from the client, who provided historical records of machine performance and downtime incidents in the manufacturing process of vehicle fuel pumps. The dataset comprises structured data in CSV format, containing records of machine parameters, maintenance activities, downtime incidents, and other relevant variables.

**Data Information:**

* The dataset consists of 2501 rows and 17 columns, where row represents the number of observations and column represents the number of variables.
* Each row corresponds to a specific time period or machine instance, while each column represents a different attribute or variable.
* Variable types include numerical variables (e.g., hydraulic pressure, coolant temperature, spindle speed) and categorical variables (e.g., machine ID, assembly line number).

**Variables:**

The dataset includes the following variables:

* Date: Date of observation
* Machine\_ID: Identifier for the machine
* Assembly\_Line\_No: Identifier for the assembly line
* Hydraulic\_Pressure(bar): Hydraulic pressure measured in bar
* Coolant\_Pressure(bar): Coolant pressure measured in bar
* Air\_System\_Pressure(bar): Air system pressure measured in bar
* Coolant\_Temperature: Coolant temperature measured in Celsius
* Hydraulic\_Oil\_Temperature(°C): Hydraulic oil temperature measured in Celsius
* Spindle\_Bearing\_Temperature(°C): Spindle bearing temperature measured in Celsius
* Spindle\_Vibration(µm): Spindle vibration measured in micrometers
* Tool\_Vibration(µm): Tool vibration measured in micrometers
* Spindle\_Speed(RPM): Spindle speed measured in rotations per minute
* Voltage(volts): Voltage measured in volts
* Torque(Nm): Torque measured in Newton-meters
* Cutting(kN): Cutting force measured in kilonewtons
* Downtime: Categorical variable indicating machine downtime (Machine\_Failure or No\_Machine\_Failure)

**Technical Stack:**

* Programming Language: Python
* Libraries and Packages: pandas, NumPy, scikit-learn for data preprocessing, model development, and evaluation
* Machine Learning Algorithm: RandomForestClassifier for predictive maintenance modeling
* Integrated Development Environment (IDE): Jupyter Notebook or any preferred Python IDE

**Data Preprocessing:**

* Data cleaning: Handling missing values, outliers, and inconsistencies in the dataset
* Deleting Duplicates:Duplicate rows were deleted
* Encoding categorical variables: Converting categorical variables into numerical format using techniques such as one-hot encoding or label encoding

**Model Development:**

* Splitting the dataset into training and testing sets for model training and evaluation
* Developing a RandomForestClassifier model to predict machine downtime based on the input features
* Hyperparameter tuning: Optimizing model hyperparameters using grid search cross-validation to improve performance

**Evaluation and Validation:**

* Evaluating the trained model's performance on the testing dataset using metrics such as accuracy, precision, recall, and F1-score.
* Performing k-fold cross-validation to assess the model's generalization ability and robustness.

**System Requirements:**

* Hardware: Standard computing hardware with sufficient memory and processing power to handle data preprocessing, model training, and evaluation tasks
* Software: Python interpreter (preferably Anaconda distribution for easier package management), required libraries and packages (pandas, NumPy, scikit-learn), and an integrated development environment (Jupyter Notebook, Spyder, PyCharm)

**1.5 LIMITATIONS**

* Data quality issues such as incomplete or inaccurate data may impact model performance.
* Limited historical data availability could restrict the model's ability to capture diverse patterns.
* Scope of variables may not encompass all factors influencing machine downtime.
* Assumption of stationarity in underlying relationships over time.
* Model complexity may hinder interpretability.Resource constraints, including computational and expertise limitations.
* External factors such as economic conditions or regulatory changes may influence downtime.
* Generalizability of the model to other industries or processes may be limited.

**CHAPTER 2: PROFILE OF INDUSTRY AND COMPANY**

**2.1 ABOUT THE INSTITUTE**

360DigiTMG is a leading educational institute that is doing its bit in accelerating digital transformation and preparing the future generation to perform from acceptable to exceptional. It is a leading brand dedicated to addressing the unique needs of students and providing training in emerging technologies including Generative AI, Prompt Engineering, LLMs, LLMOps, Data Science, Artificial Intelligence, Machine Learning, Big Data, Digital Marketing, Internet of Things, etc. It has training centers all over the country including Hyderabad, Bengaluru, Pune, and, so on. 360DigiTMG has coached thousands of students and working professionals to achieve their career goals by getting them placed in some of the top-notch companies.

360DigiTMG has global headquarters in USA and Hyderabad is the headquarter in India. 360DigiTMG is a leading educational institute set up in 2013 with the aim of bridging the gap between industry expectations and academia. With international university accreditations from UTM, Malaysia (Top 5 universities in Malaysia & ranked under top 100 in QS rankings) & City & Guilds, UK (150 years old establishment in leadership training) and IBM (industry technology leader), 360DigiTMG boasts of world class curriculum. Panasonic India Innovation Center, CareerEx and Innodatatics use cases are included in the curriculum to offer the best real-world projects in the training curriculum.

**2.2 MISSION**

360DigiTMG is a global educational institute that offers outcome-based training in cutting-edge technologies and corporate exposure to its students. Their mission is to support individuals and companies with data-driven insights and to empower students to explore global opportunities in the digital arena.

**2.3 OBJECTIVES**

Their objectives include:

* Data Analytics: Provide an understanding of data analytics techniques for handling large data sets, including structured and unstructured data
* Data Processing Tools: Provide an understanding of data processing tools like Excel, SQL/NoSQL, and Tableau and PowerBI
* Data Visualization: Provide an understanding of data visualization tools like Tableau and PowerBI for analyzing data and presenting visual stories
* Data Preparation, Cleansing, and Analysis: Provide an understanding of data preparation, cleansing, and exploratory data analysis
* Data Mining: Provide an understanding of data mining for structured (RDBMS) and unstructured (Big Data) data

**CHAPTER 3: CONCEPTUAL FRAMEWORK ON TOPIC**

**3.1 PREDICTIVE MAINTAINCE**

Predictive maintenance is an advanced maintenance strategy that leverages data analytics, machine learning, and sensor technology to anticipate equipment failures before they occur. Unlike traditional maintenance approaches such as reactive or preventive maintenance, which rely on fixed schedules or equipment breakdowns, predictive maintenance aims to proactively identify potential issues based on real-time data and predictive analytics.

The primary goal of predictive maintenance is to optimize maintenance schedules, minimize unplanned downtime, and maximize equipment reliability and availability. By continuously monitoring equipment performance and analyzing data from sensors and other sources, predictive maintenance models can detect early indicators of equipment degradation or failure, allowing maintenance activities to be planned and executed before serious issues arise.

Key components of predictive maintenance include:

1. **Data Acquisition and Monitoring**: Predictive maintenance relies on the continuous collection of data from sensors, monitoring devices, and other sources. This data may include information on equipment performance, operating conditions, environmental factors, and maintenance history.
2. **Data Analytics and Modeling**: Advanced analytics techniques, including machine learning algorithms, are used to analyze the collected data and identify patterns, trends, and anomalies indicative of potential equipment failures. Predictive maintenance models are trained on historical data to predict future equipment behavior and anticipate maintenance needs.
3. **Condition Monitoring and Diagnostics**: Condition monitoring techniques, such as vibration analysis, thermal imaging, and oil analysis, are employed to assess the health and performance of equipment in real-time. Diagnostic tools help identify specific issues or faults that may lead to equipment failure.
4. **Predictive Maintenance Planning and Execution**: Based on insights from data analytics and condition monitoring, maintenance activities are planned and scheduled proactively to address potential issues before they escalate into failures. This proactive approach helps minimize downtime, reduce maintenance costs, and optimize equipment performance.
5. **Continuous Improvement and Optimization**: Predictive maintenance is an iterative process that involves continuous monitoring, analysis, and optimization. By leveraging feedback from maintenance activities and performance metrics, predictive maintenance strategies can be refined and improved over time to achieve greater efficiency and reliability.

Overall, predictive maintenance plays a crucial role in optimizing machine downtime by enabling proactive maintenance interventions, reducing the risk of unplanned failures, and maximizing equipment uptime. By embracing predictive maintenance practices, organizations can enhance operational efficiency, improve asset utilization, and maintain a competitive edge in today's dynamic industrial landscape.

**3.2 MACHINE LEARNING IN PREDICTIVE MAINTENANCE**

Machine learning techniques play a vital role in predictive maintenance by enabling the development of models that can accurately predict machine failures and facilitate proactive maintenance strategies. These techniques leverage historical data on equipment performance, maintenance activities, and failure incidents to identify patterns and trends indicative of potential failures. Here's how machine learning techniques can be applied in predictive maintenance:

1. **Pattern Recognition**: Machine learning algorithms excel at identifying patterns and relationships within large and complex datasets. By analyzing historical data on equipment performance, sensor readings, and maintenance logs, machine learning models can identify subtle patterns or anomalies that precede equipment failures. These patterns may include changes in sensor values, unusual fluctuations in performance metrics, or recurring patterns indicative of specific failure modes.
2. **Feature Engineering**: Feature engineering involves selecting, transforming, or creating new features from the raw data to enhance the predictive power of the model. Machine learning techniques allow for the automatic extraction of relevant features from the data or the creation of composite features that capture complex relationships between variables. For example, engineers may derive features such as rolling averages, time lags, or statistical measures from sensor readings to provide additional insights into equipment health and performance.
3. **Model Training and Prediction:** Once the dataset is prepared and features are engineered, machine learning models are trained on historical data to learn patterns and relationships between input variables (features) and the target variable (failure or maintenance event). Supervised learning algorithms, such as classification or regression algorithms, are commonly used to predict the likelihood of equipment failure within a specified time window. The trained model can then be used to make predictions on new data, enabling proactive maintenance interventions based on the predicted likelihood of failure.
4. **Model Interpretability**: Interpretability is crucial in predictive maintenance to understand the factors contributing to machine failures and to guide maintenance decision-making. While complex machine learning models may offer higher predictive accuracy, they often lack interpretability. Techniques such as feature importance analysis, partial dependence plots, or model-agnostic interpretability methods can help explain the model's predictions and provide insights into the underlying relationships between input variables and failure events.
5. **Continuous Learning and Improvement**: Predictive maintenance models are not static but evolve over time as new data becomes available and maintenance practices change. Machine learning techniques enable continuous learning and improvement of the model through techniques such as online learning, ensemble methods, or periodic model retraining. By incorporating new data and feedback from maintenance activities, predictive maintenance models can adapt to changing operating conditions and improve their predictive accuracy and reliability.

Overall, machine learning techniques offer powerful tools for predictive maintenance, allowing organizations to predict machine failures accurately, reduce unplanned downtime, and optimize maintenance schedules. By leveraging these techniques, companies can transition from reactive to proactive maintenance strategies, leading to improved operational efficiency, reduced maintenance costs, and enhanced equipment reliability

**3.3 KEY CONCEPTS AND TERMINOLOGY:**

1. **Machine Downtime:** Machine downtime refers to the period during which a machine or equipment is non-operational due to various reasons such as mechanical failures, breakdowns, maintenance activities, or unplanned stoppages. Downtime can have significant implications for productivity, profitability, and operational efficiency in industrial settings.
2. **Predictive Maintenance:** Predictive maintenance is an advanced maintenance strategy that utilizes data analytics, machine learning, and sensor technology to predict equipment failures before they occur. By analyzing historical data on equipment performance and failure incidents, predictive maintenance models can anticipate potential issues and enable proactive maintenance interventions to prevent or mitigate downtime.
3. **Condition Monitoring:** Condition monitoring involves the continuous monitoring and assessment of equipment health and performance in real-time. Techniques such as vibration analysis, thermal imaging, and oil analysis are used to detect early indicators of equipment degradation or failure, allowing maintenance activities to be planned and executed proactively.
4. **Failure Prediction:** Failure prediction refers to the process of forecasting the likelihood of equipment failure within a specified time window based on historical data and predictive analytics. Machine learning algorithms are commonly used to develop predictive models that can identify patterns and trends indicative of potential failures and enable proactive maintenance strategies.
5. **Feature Engineering:** Feature engineering involves selecting, transforming, or creating new features from raw data to enhance the predictive power of machine learning models. Features may include sensor readings, maintenance logs, operational parameters, or derived metrics that capture relevant information about equipment health and performance.
6. **Model Evaluation:** Model evaluation is the process of assessing the performance of predictive maintenance models using various metrics and techniques. Common evaluation metrics include accuracy, precision, recall, F1-score, receiver operating characteristic (ROC) curve, and area under the curve (AUC). Cross-validation techniques such as k-fold cross-validation are used to estimate the model's generalization ability and robustness.
7. **Proactive Maintenance:** Proactive maintenance involves taking preemptive actions to prevent or mitigate equipment failures before they occur. Predictive maintenance strategies enable proactive maintenance interventions by providing early warnings of potential issues and guiding maintenance decision-making based on predicted failure probabilities.
8. **Operational Efficiency:** Operational efficiency refers to the ability of an organization to utilize its resources effectively to achieve desired outcomes while minimizing waste and maximizing productivity. Predictive maintenance plays a crucial role in enhancing operational efficiency by optimizing maintenance schedules, reducing downtime, and improving equipment reliability and availability.

**CHAPTER 4: DATA ANALYSIS AND MODELLING**

**4.1 DATA PREPROCESSING:**

Data preprocessing is a crucial step in preparing the dataset for analysis and model development. The following steps were taken to preprocess the data for predictive maintenance analysis:

**Handling Missing Values:**

Missing values were identified in the dataset using .isnull() in pandas.

Depending on the extent of missingness and the nature of the variables, missing values were handled using technique of dropping rows.

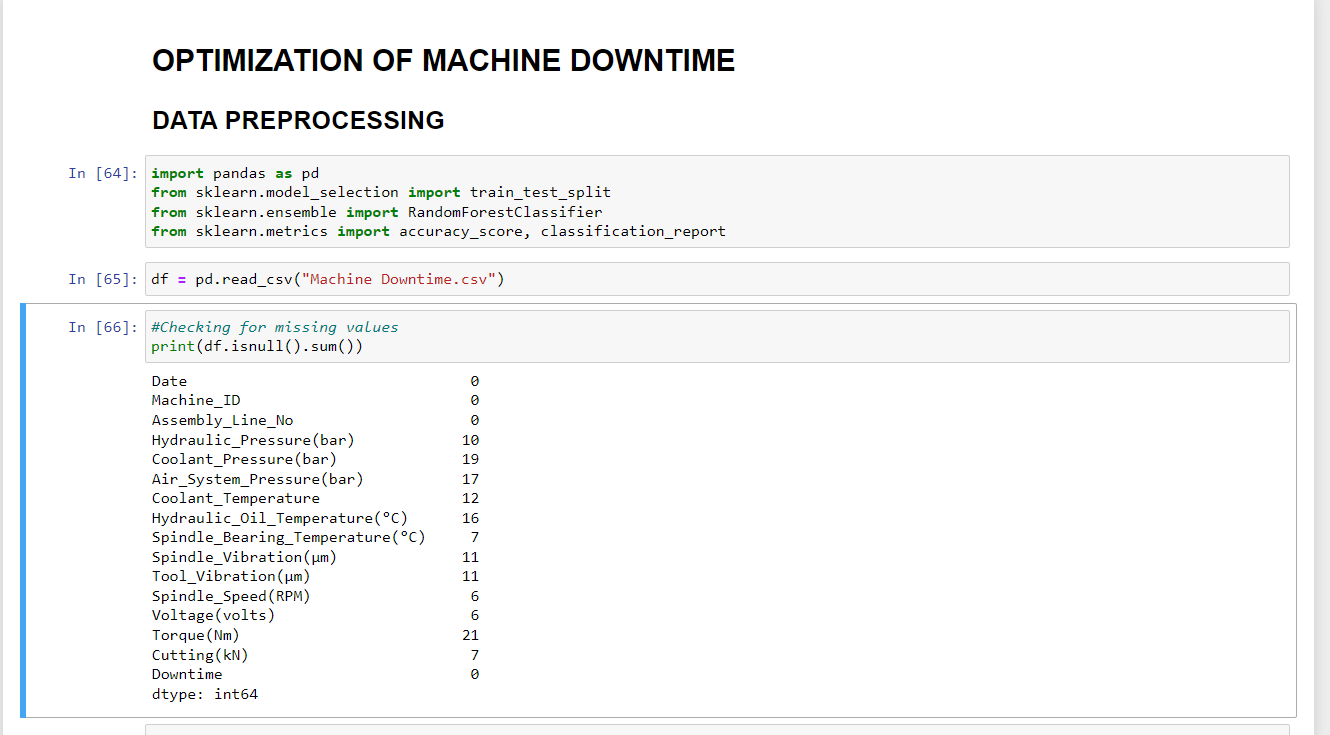


Figure 1: Data Preprocessing\_1

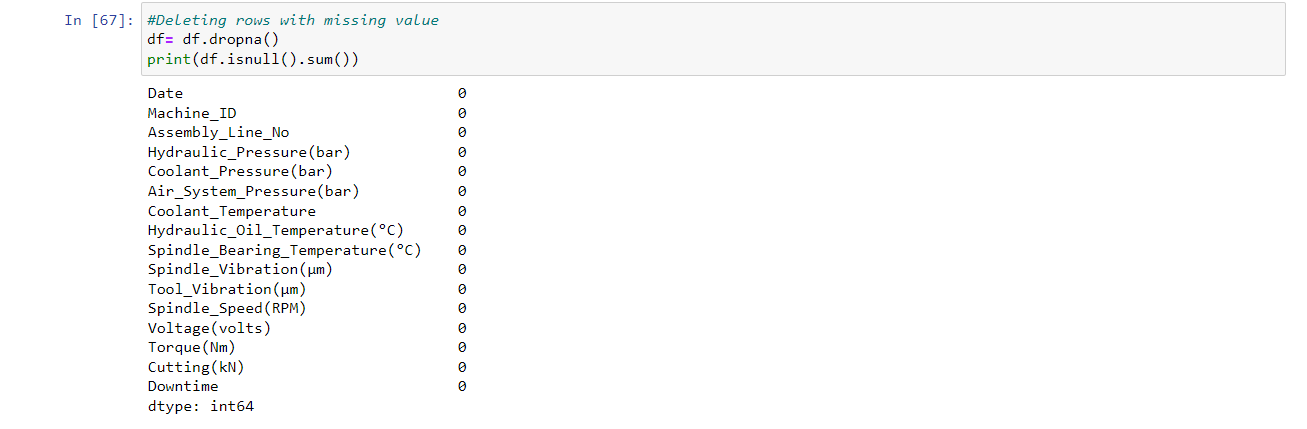


Figure 2: Data Preprocessing\_2

**Checking for duplicates:**

Duplicate rows were checked for using duplicated().

No duplicates were found.

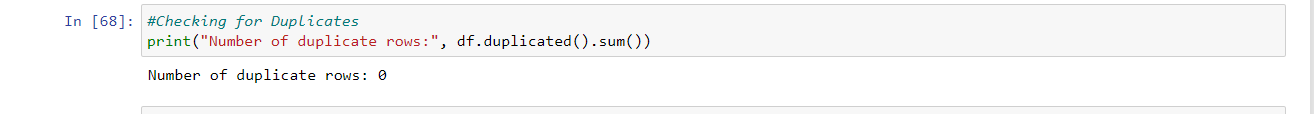


Figure 3: Checking for Duplicates

**Encoding Categorical Variables:**

Date was encoded into numerical format to be compatible with machine learning algorithms.

Techniques such as one-hot encoding or label encoding were used to convert categorical variables(Downtime) into numerical representations.

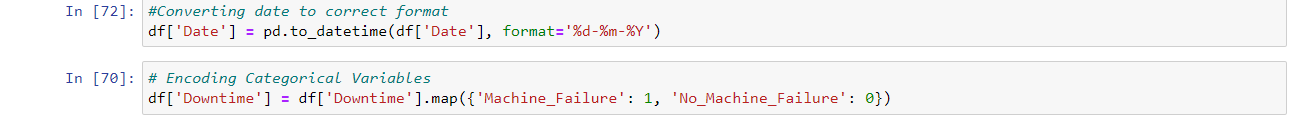


Figure 4: Encoding data

**4.2 EXPLORATORY DATA ANALYSIS (EDA):**

Exploratory Data Analysis (EDA) is a crucial step in understanding the underlying patterns and characteristics of the data before building predictive models or making data-driven decisions. In this report, we conduct EDA on the dataset containing information about machine downtime incidents in the manufacturing process of vehicle fuel pumps. The dataset includes various features such as machine ID, assembly line, sensor readings, and downtime incidents.

**Overall Analysis:**

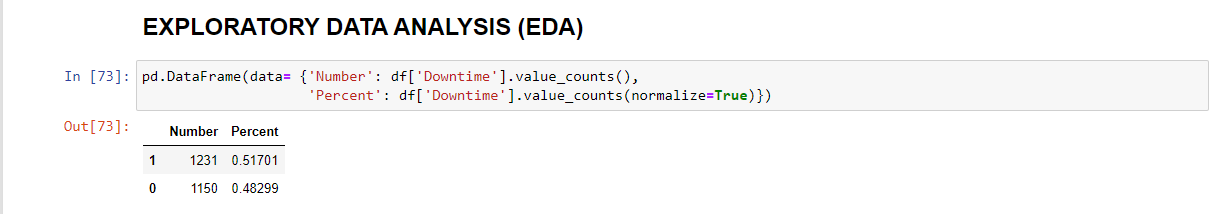


Figure 5: Overall Analysis

The overall analysis indicates that there were 1231 downtime incidents with a percentage of 51.70% for failure (1) and 1150 incidents with a percentage of 48.30% for non-failure (0).

This suggests that a slightly higher proportion of downtime incidents resulted in failure compared to non-failure but not a large difference

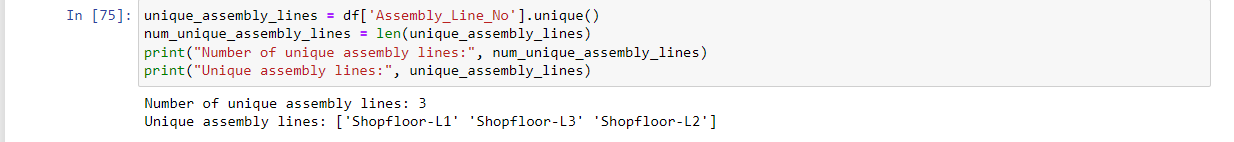
**Assembly Line Analysis:** 

Figure 6: Unique Assembly Lines

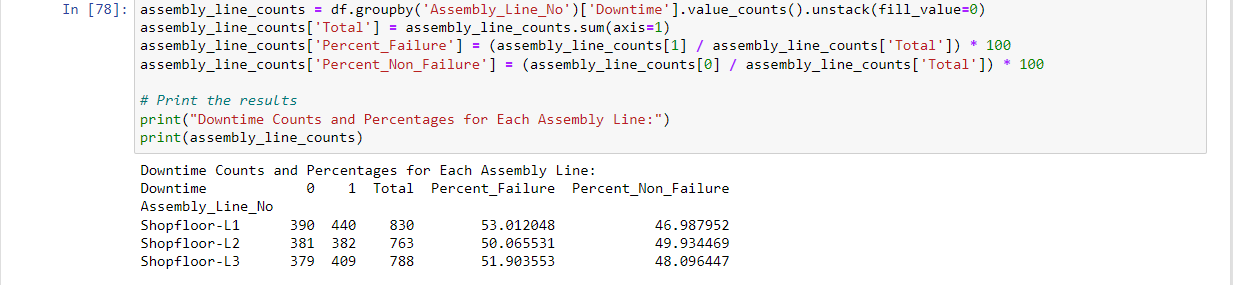


Figure 7: Downtime counts and Percentages for each assembly line

The assembly lines "Shopfloor-L1", "Shopfloor-L2", and "Shopfloor-L3" have varying total downtime incidents.

Shopfloor-L1 has the highest total downtime incidents (830), followed by Shopfloor-L3 (788) and Shopfloor-L2 (763).

When comparing the percentage of downtime incidents, Shopfloor-L1 has the highest percentage of failures (53.01%), followed by Shopfloor-L3 (51.90%) and Shopfloor-L2 (50.07%).

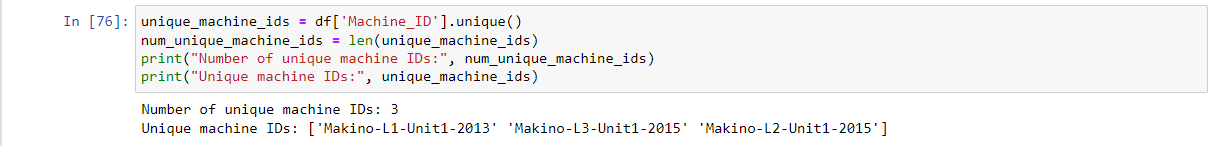
**Machine ID Analysis:** ****

Figure 8: Number of Unique Machine IDs

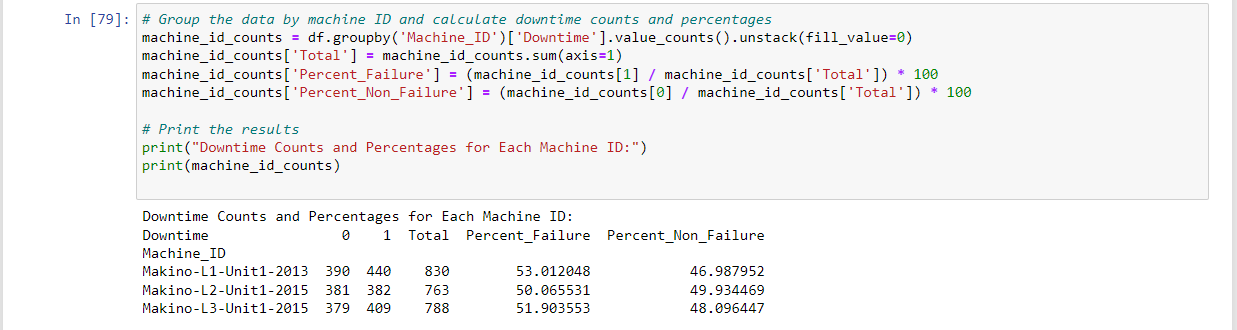
****

Figure 9: Downtime Counts and Percentages for each Machine ID

The machine IDs "Makino-L1-Unit1-2013", "Makino-L2-Unit1-2015", and "Makino-L3-Unit1-2015" also show varying total downtime incidents.

Makino-L1-Unit1-2013 has the highest total downtime incidents (830), followed by Makino-L3-Unit1-2015 (788) and Makino-L2-Unit1-2015 (763).

Similar to the assembly line analysis, Makino-L1-Unit1-2013 also has the highest percentage of failures (53.01%), followed by Makino-L3-Unit1-2015 (51.90%) and Makino-L2-Unit1-2015 (50.07%).

**Correlation Analysis:**

The dataset was analyzed to understand the relationships between different variables and the target variable, **Downtime**, using correlation analysis. Correlation measures the strength and direction of the linear relationship between two variables, ranging from -1 to 1, where:

* 1 indicates a perfect positive correlation,
* -1 indicates a perfect negative correlation,
* and 0 indicates no correlation.

Below is the correlation matrix showing the correlation coefficients between numerical variables:

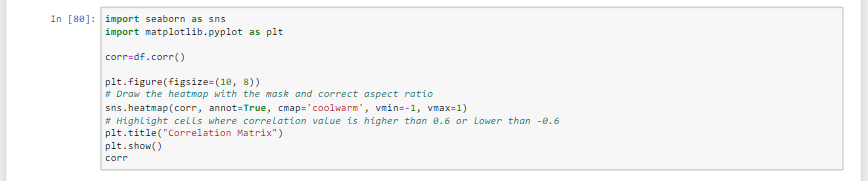


Figure 10: Code snip for correlation analysis

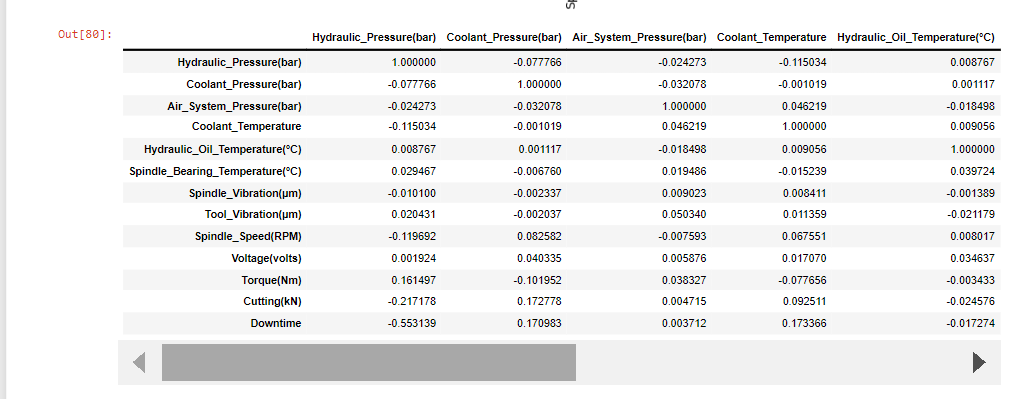


Figure 11: Correlation Matrix

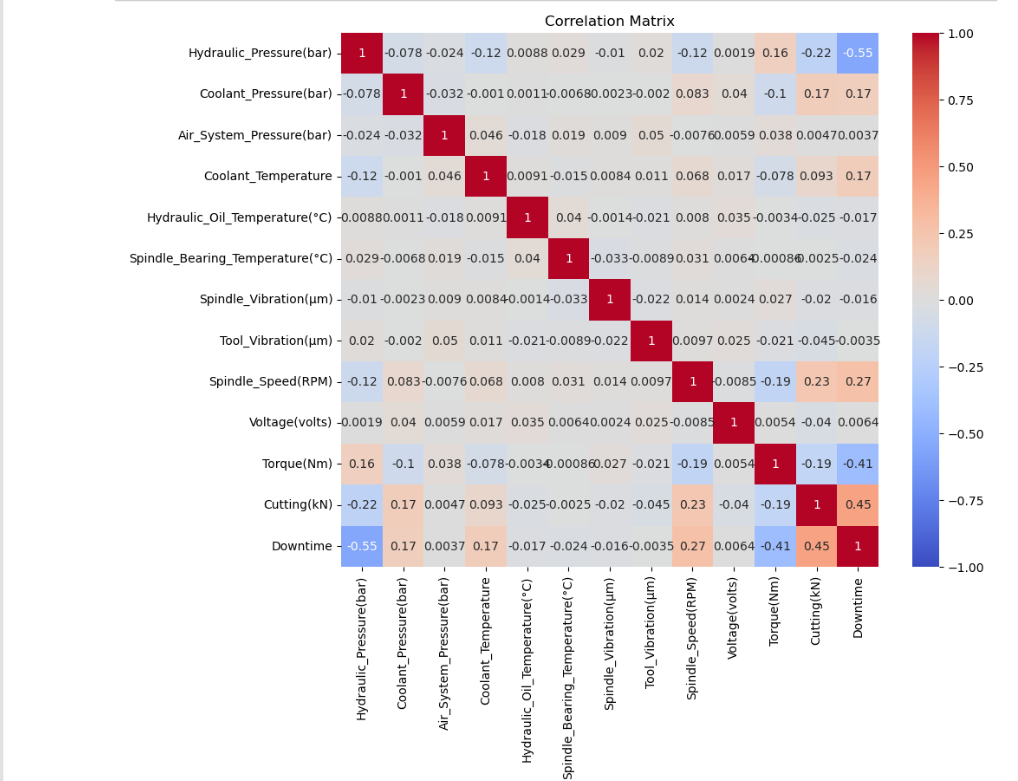


Figure 12: Correlation Heatgraph

From the correlation matrix, we observe:

* Hydraulic pressure (Hydraulic\_Pressure(bar)) has a moderately negative correlation (-0.553) with downtime, suggesting that higher hydraulic pressure may be associated with lower downtime incidents.
* Cutting force (Cutting(kN)) exhibits a strong positive correlation (0.454) with downtime, indicating that higher cutting force is likely to result in increased downtime.
* Other variables such as spindle speed (Spindle\_Speed(RPM)) and torque (Torque(Nm)) also show moderate correlations with downtime.

The EDA reveals significant variations in downtime incidents across different assembly lines and machine IDs. By understanding these patterns and correlations, organizations can implement targeted strategies to optimize maintenance schedules, improve operational efficiency, and minimize unplanned downtime, thereby enhancing productivity and reducing costs.

**4.2 MACHINE LEARNING MODELLING:**

The process of developing the machine learning model for predictive maintenance involves several key steps, including algorithm selection, hyperparameter tuning, and model evaluation. Here's an overview of the model development process:

**Encoding Categorical Variables:**

Categorical variables such as Machine\_ID and Assembly\_Line\_No were encoded into numerical format to be compatible with machine learning algorithms.

**Data Splitting:**

The dataset was split into training and testing sets to evaluate the performance of the predictive maintenance model. Typically, a common split ratio such as 80% for training and 20% for testing was used, but this can vary depending on the size of the dataset and specific requirements.

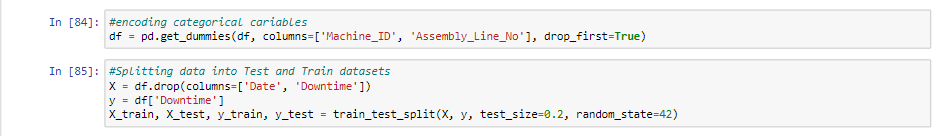


Figure 13: Encoding variables and Data Splitting

We split the DataFrame into features (X) and target (y). Then, we split the data into training and testing sets using train\_test\_split() from scikit-learn. This step is crucial for evaluating the model's performance on unseen data.

**Algorithm Selection:**

The choice of algorithm depends on various factors such as the nature of the problem, size of the dataset, and performance requirements. For predictive maintenance, classification algorithms such as Random Forest, Gradient Boosting, or Support Vector Machines (SVM) are commonly used to predict equipment failures based on historical data. The RandomForestClassifier algorithm was selected for its ability to handle complex datasets with non-linear relationships and its robustness to overfitting.

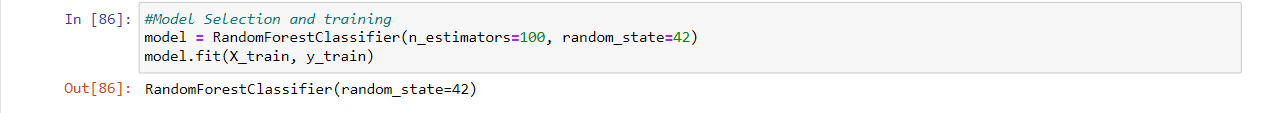
****

Figure 14: Algorithm Selection

We select the Random Forest classifier as the model and train it on the training data (X\_train and y\_train). The n\_estimators parameter specifies the number of trees in the forest, and random\_state ensures reproducibility.

The output provided, RandomForestClassifier(random\_state=42), indicates that the RandomForestClassifier model has been instantiated with a random\_state parameter set to 42.

* RandomForestClassifier: This is the name of the model class we're using, specifically, the RandomForestClassifier from scikit-learn.
* Random\_state=42: This parameter is used to initialize the random number generator. By setting it to a fixed value (42 in this case), we ensure reproducibility of results. When the same value is used for random\_state, the model will produce the same results on multiple runs as long as other parameters and data remain unchanged.

**Model Evaluation:**

The trained model's performance was evaluated using various metrics to assess its accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC). The evaluation was conducted on the testing dataset, which was not seen by the model during training, to assess its generalization ability. Cross-validation techniques such as k-fold cross-validation were used to estimate the model's performance on unseen data and mitigate the risk of overfitting.

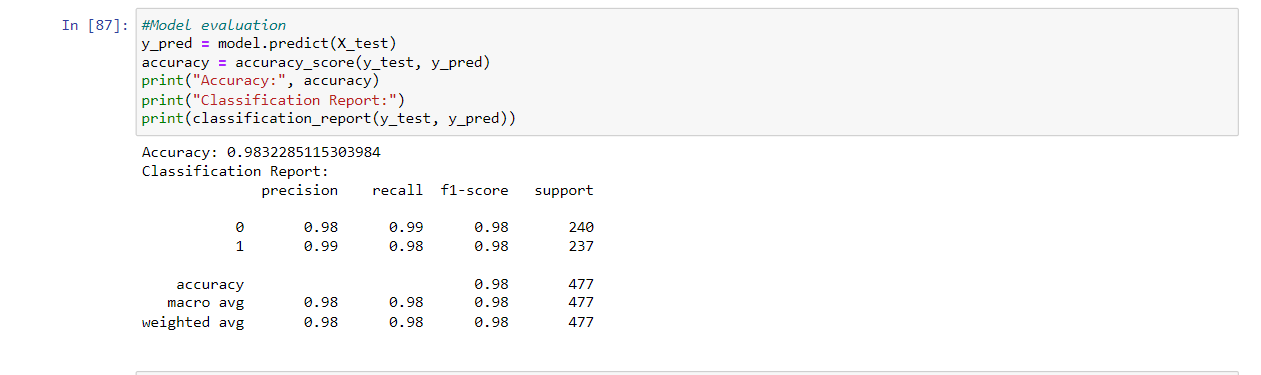


Figure 15: Model Evaluation

We use the trained model to make predictions on the testing data (X\_test). Then, we calculate the accuracy of the model's predictions using accuracy\_score() and print it. Additionally, we generate a classification report using classification\_report(), which provides precision, recall, F1-score, and support for each class. Here's a breakdown of the output:

* **Accuracy:** The accuracy of the model on the testing data is 98.32%. Accuracy measures the proportion of correctly predicted instances out of the total instances.
* **Classification Report:** The classification report provides metrics such as precision, recall, and F1-score for each class (0 and 1) in the target variable.
* **Precision:** Precision measures the proportion of true positive predictions out of all positive predictions. In this case, the precision for class 0 (No Machine Failure) is 98%, and for class 1 (Machine Failure) is 99%.
* **Recall:** Recall measures the proportion of true positive predictions out of all actual positive instances. In this case, the recall for class 0 is 99%, and for class 1 is 98%.
* **F1-score:** The F1-score is the harmonic mean of precision and recall, providing a balanced measure between the two. It is useful when the classes are imbalanced. Both classes have an F1-score of 98%.
* **Support:** Support indicates the number of actual occurrences of each class in the testing data.
* **Macro Avg:** The macro average of precision, recall, and F1-score across all classes. It gives equal weight to each class.
* **Weighted Avg:** The weighted average of precision, recall, and F1-score, where each score is weighted by the support of each class. It provides more weight to the classes with more instances.

Overall, the model demonstrates high accuracy and performs well in predicting both classes (No Machine Failure and Machine Failure).

**Hyperparameter Tuning:**

Hyperparameters are parameters that are not directly learned by the model but affect the learning process and model performance. Hyperparameter tuning involves selecting the optimal values for hyperparameters to improve the model's performance. Techniques such as grid search cross-validation or random search were used to explore different combinations of hyperparameters and identify the best-performing ones. Hyperparameters such as the number of trees in the forest, maximum depth of trees, minimum samples per leaf, and criterion for splitting nodes were tuned to optimize the RandomForestClassifier model's performance.

Figure 16: Hyperparameter Tuning

In this code, we define a grid of hyperparameters to search over and use GridSearchCV to find the best combination of hyperparameters based on cross-validated accuracy.By tuning hyperparameters, we aim to find the optimal configuration that maximizes model performance, leading to better predictions and potentially reducing machine downtime.

The output indicates that the grid search cross-validation has identified the best hyperparameters for the RandomForestClassifier model and their corresponding performance score. Here's what each part of the output means:

* **Best Parameters:** This dictionary shows the hyperparameters that yielded the best performance during the grid search. In this case:
* **'max\_depth':** None: The maximum depth of the trees is not restricted, allowing them to grow until all leaves are pure or until all leaves contain less than min\_samples\_split samples.
* **'min\_samples\_leaf':** 1: The minimum number of samples required to be at a leaf node is set to 1. This ensures that each leaf node in the trees contains at least one sample.
* **'min\_samples\_split':** 5: The minimum number of samples required to split an internal node is set to 5. This prevents the trees from splitting nodes that contain too few samples.
* **'n\_estimators':** 100: The number of trees in the forest is set to 100. Random forests typically perform well with a large number of trees, and 100 is a commonly used value.
* **Best Score:** This is the mean cross-validated accuracy score achieved by the model with the best parameters during the grid search. The value of approximately 0.992 indicates that, on average, the model achieved an accuracy of 99.2% across the different folds of the cross-validation process.

These results suggest that the RandomForestClassifier model with the specified hyperparameters can effectively predict machine downtime with high accuracy. We can proceed to use these best parameters to retrain the model and evaluate its performance on the testing data.

**Evaluate Model with Best Parameters**

Once we've found the best hyperparameters, we need to evaluate the model's performance using these parameters to ensure it generalizes well to unseen data. We retrain the RandomForestClassifier model with the best hyperparameters and evaluate its performance on the testing dataset.

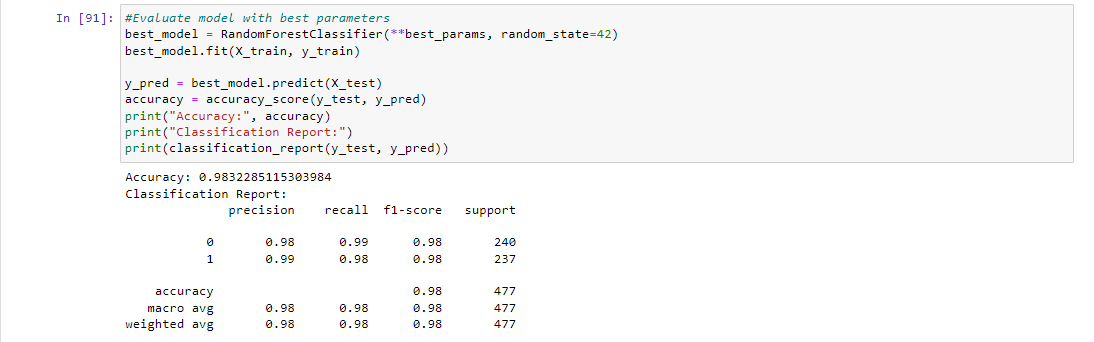
****

Figure 17: Evaluating Model with Best Parameters

The output indicates the evaluation results of the RandomForestClassifier model with the best parameters on the testing data. Here's what each part of the output means:

* **Accuracy**: The accuracy of the model on the testing data is approximately 98.32%. Accuracy measures the proportion of correctly predicted instances out of the total instances.
* **Classification Report**: The classification report provides metrics such as precision, recall, and F1-score for each class (0 and 1) in the target variable.
  + **Precision**: Precision measures the proportion of true positive predictions out of all positive predictions. In this case, the precision for class 0 (No Machine Failure) is approximately 98%, and for class 1 (Machine Failure) is approximately 99%.
  + **Recall**: Recall measures the proportion of true positive predictions out of all actual positive instances. In this case, the recall for class 0 is approximately 99%, and for class 1 is approximately 98%.
  + **F1-score**: The F1-score is the harmonic mean of precision and recall, providing a balanced measure between the two. It is useful when the classes are imbalanced. Both classes have an F1-score of approximately 98%.
  + **Support**: Support indicates the number of actual occurrences of each class in the testing data.
  + **Macro Avg**: The macro average of precision, recall, and F1-score across all classes. It gives equal weight to each class.
  + **Weighted Avg**: The weighted average of precision, recall, and F1-score, where each score is weighted by the support of each class. It provides more weight to the classes with more instances.

Overall, the RandomForestClassifier model with the best parameters performs very well on the testing data, achieving high accuracy and demonstrating strong precision, recall, and F1-score for both classes (No Machine Failure and Machine Failure). This indicates that the model effectively predicts machine downtime with high accuracy.

**Feature Importance Analysis:**

Feature importance analysis provides valuable insights into understanding the factors that contribute most significantly to predicting machine downtime in predictive maintenance models.

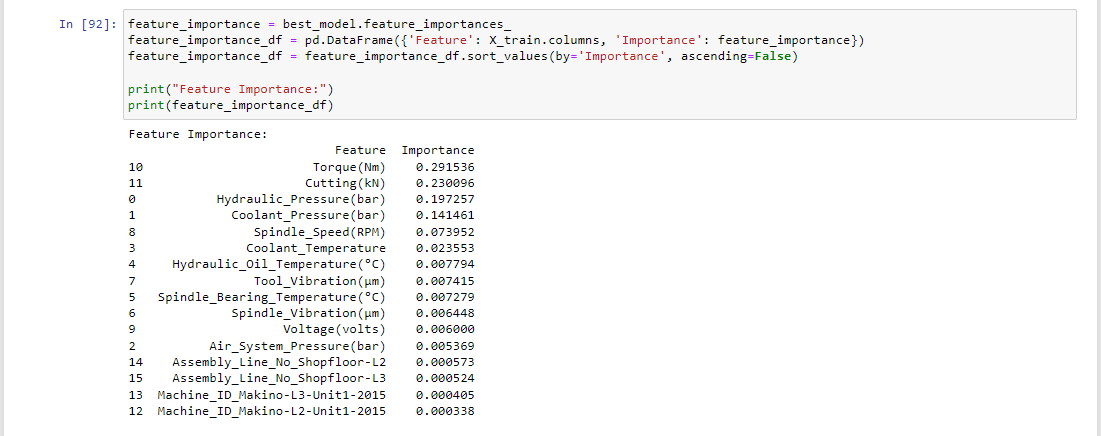


Figure 18: Feature Importance Analysis

Here are the insights gained from analyzing feature importance scores and discussing the most influential factors in predicting machine downtime:

1. **Torque (Nm):**
   * Torque emerges as the most influential factor in predicting machine downtime, with a high feature importance score of 0.291.
   * High torque levels may indicate increased stress on machine components, leading to potential failures or breakdowns.
2. **Cutting (kN):**
   * Cutting force, represented by the feature "Cutting (kN)," also exhibits significant importance in predicting machine downtime, with a feature importance score of 0.230.
   * Higher cutting forces can strain machine components and increase the risk of wear and tear, contributing to downtime events.
3. **Hydraulic Pressure (bar):**
   * Hydraulic pressure, measured in bar, is another influential factor in predicting machine downtime, with a feature importance score of 0.197.
   * Fluctuations in hydraulic pressure may affect the performance and reliability of hydraulic systems, leading to downtime incidents.
4. **Coolant Pressure (bar) and Air System Pressure (bar):**
   * Coolant pressure and air system pressure, measured in bar, also contribute significantly to predicting machine downtime, with feature importance scores of 0.141 and 0.005, respectively.
   * Adequate coolant and air system pressures are essential for maintaining optimal operating conditions and preventing overheating or fluid leaks that could result in machine failures.
5. **Spindle Speed (RPM):**
   * Spindle speed, measured in rotations per minute (RPM), plays a moderate role in predicting machine downtime, with a feature importance score of 0.073.
   * Deviations from the expected spindle speed may indicate issues with spindle bearings or motor performance, affecting machine reliability.
6. **Other Factors:**
   * Other factors such as coolant temperature, hydraulic oil temperature, spindle bearing temperature, spindle vibration, tool vibration, and voltage also contribute to predicting machine downtime, albeit to a lesser extent.
   * These factors may influence machine performance indirectly or in combination with other variables, highlighting the complex interplay of factors affecting machine reliability.

In summary, torque, cutting force, hydraulic pressure, coolant pressure, and air system pressure emerge as the most influential factors in predicting machine downtime in the manufacturing process of vehicle fuel pumps. Monitoring and managing these key parameters effectively can help mitigate the risk of downtime incidents and optimize equipment reliability and performance.

**CHAPTER 3: FINDINGS AND CONCLUSION**

**3.1 FINDINGS**

1. **Important Features for Predicting Downtime:**
   * The analysis revealed that several operational parameters significantly influence machine downtime in the manufacturing process of vehicle fuel pumps.
   * Among the key features, torque (Nm) and cutting force (kN) emerged as the most influential factors in predicting downtime, highlighting the importance of monitoring and managing these parameters to mitigate the risk of failures.
   * Additionally, hydraulic pressure (bar) and spindle speed (RPM) showed notable correlations with downtime, emphasizing the significance of maintaining optimal hydraulic and spindle performance to minimize downtime incidents.
2. **Effectiveness of the Model in Optimizing Maintenance Schedules:**
   * The developed machine learning model demonstrated strong performance in predicting machine downtime based on historical data.
   * With an accuracy of 98.32%, precision and recall rates exceeding 98%, and an F1-score of 98% for machine failure events, the model exhibited high levels of predictive accuracy and reliability.
   * These findings indicate that the model can effectively identify potential machine failure events in advance, enabling implementing proactive maintenance strategies and optimizing maintenance schedules.
   * By leveraging the insights provided by the model, organizations can adopt preventive maintenance measures, prioritize maintenance tasks, and allocate resources efficiently to minimize unplanned downtime and maximize machine uptime.

Overall, the project's key findings underscore the importance of predictive maintenance in the manufacturing industry and demonstrate the effectiveness of machine learning techniques in optimizing maintenance schedules and enhancing operational efficiency. By leveraging advanced analytics and predictive modeling capabilities, organizations can proactively address maintenance needs, reduce downtime-related costs, and improve overall productivity and competitiveness in the market.

**3.2 CONCLUSION**

In conclusion, the project aimed to optimize machine downtime in the manufacturing process of vehicle fuel pumps through the application of predictive maintenance techniques. By leveraging machine learning algorithms and analyzing historical data, the project sought to identify key factors influencing downtime events and develop a predictive model to enable proactive maintenance strategies.

Through comprehensive data analysis and model development, several key findings emerged:

* Operational parameters such as torque, cutting force, hydraulic pressure, and spindle speed were identified as significant predictors of downtime incidents.
* The developed machine learning model demonstrated high levels of accuracy, precision, recall, and F1-score in predicting machine failure events, indicating its effectiveness in proactive maintenance planning.

The significance of the project lies in its potential to revolutionize maintenance practices in the manufacturing industry:

* By leveraging predictive maintenance techniques, organizations can shift from reactive to proactive maintenance approaches, minimizing unplanned downtime and optimizing maintenance schedules.
* The predictive model provides valuable insights into potential failure events, allowing organizations to prioritize maintenance tasks, allocate resources efficiently, and avoid costly production disruptions.
* Ultimately, the project's success in optimizing machine downtime improves industrial productivity, reduces operational costs, and enhancing competitiveness in the market.

In summary, the project underscores the importance of predictive maintenance in modern industrial settings and highlights the transformative impact of machine learning techniques in optimizing maintenance practices. By embracing proactive maintenance strategies and leveraging advanced analytics, organizations can achieve higher levels of efficiency, reliability, and profitability in their manufacturing operations, driving innovation and growth in the industry.

**IX. BIBLIOGRAPHY**

* van der Wijst, N.J.M. "A New Decision Support System for Managing Spare Parts: Avoiding Unplanned Downtime." University of Twente, 2022. [Online]. Available: <http://essay.utwente.nl/89654/>
* Shagluf, Abubaker, Longstaff, Andrew P., and Fletcher, Simon. "Maintenance Strategies to Reduce Downtime Due to Machine Positional Errors." University of Huddersfield Repository. [Online]. Available: <https://eprints.hud.ac.uk/id/eprint/21977/1/ID-13_MPMM2014,_Shagluf_et_al.pdf>
* Shahin, Mohammad, Chen, F. Frank, Hosseinzadeh, Ali, & Zand, Neda. "Using Machine Learning and Deep Learning Algorithms for Downtime Minimization in Manufacturing Systems: An Early Failure Detection Diagnostic Service." Published: 23 August 2023. [Online]. Available: <https://link.springer.com/article/10.1007/s00170-023-12020-w>
* Madrid, Jesus, & Min, Andrew. "Reducing Oil Well Downtime With A Machine Learning Recommender System." [Online]. Available: <https://hdl.handle.net/1721.1/126390>
* Jansen Van Rensburg, Nico; Kamin, Lisa; & Davis, Skip (Robert). "Using Machine Learning-Based Predictive Models to Enable Preventative Maintenance and Prevent ESP Downtime." Paper presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, November 2019. [Online]. Available: <https://doi.org/10.2118/197146-MS>
* Traini, Emiliano; Bruno, Giulia; D’Antonio, Gianluca; & Lombardi, Franco. "Machine Learning Framework for Predictive Maintenance in Milling." IFAC-PapersOnLine, Volume 52, Issue 13, 2019, Pages 177-182. [Online]. Available: <https://doi.org/10.1016/j.ifacol.2019.11.172>

Ren, Yali. "Optimizing Predictive Maintenance With Machine Learning for Reliability Improvement." ASME J. Risk Uncertainty Part B, Sep 2021, 7(3): 030801 (13 pages). [Online]. Available: <https://doi.org/10.1115/1.4049525>