**Capstone Project**

**Analyzing Machine Downtime**

**ST. CLAIR COLLEGE**

GROUP MEMBERS

Mirenkumar Patel(0782036)

Zeel Mendapara(0790824)

Nirmit Pareliya(0785084)

Miloni Patel(0788500)

Harsh Patel(0791820)

PROF: JOHN GERASSIMOU

**Abstract**

Machine downtime, which refers to the time when a machine or piece of equipment is not in use due to planned maintenance, repairs, or unforeseen breakdowns, is a significant issue that firms in numerous sectors must deal with. Equipment failures, unplanned maintenance, human mistake, and poor training are just a few of the many factors that can lead to downtime. Businesses may experience serious effects from machine downtime, such as lost production time, missed deadlines, lowered product quality, and higher prices.

Companies can use a variety of tactics and devices to lessen the consequences of machine downtime. Implementing a preventative maintenance programme with frequent inspections, maintenance, and repairs is a popular strategy. Utilizing systems that utilize analytics and data to forecast when a machine is likely to break and plan maintenance appropriately is another tactic. Recent technological developments have made it feasible to utilize data and analytics to pinpoint the underlying causes of downtime and take preventative action.

Businesses must comprehend the reasons for machine downtime and apply efficient tactics and technologies to mitigate the consequences if they want to increase overall efficiency and lessen the impact of machine downtime. Businesses may use data and analytics to reduce downtime and enhance their operations by taking pre-emptive actions.

**Acknowledgements**

We want to express our heartfelt appreciation to FreePoint Inc. for giving us the chance to collaborate on the project. This project has been a fun and difficult opportunity for us to put our data analysis, predictive modelling, and machine learning talents to use to solve actual business issues.

We would like to sincerely thank the project supervisor for their guidance, support, and encouragement throughout the project. Their expertise and knowledge have been invaluable in helping us to understand the complex concepts related to machine downtime analysis.

Additionally, we want to express our appreciation to the team for working well together. To complete the project and get the desired results, the team's joint effort was essential.

Finally, we want to thank everyone who contributed to this project, including the many stakeholders. They have played a significant role in determining the project's course and guaranteeing its success through their ideas, feedback, and suggestions.

**Table of Contents**

1. **INTRODUCTION…………………………………………………….............................5**
2. **DATASET………...….………………………………………………...………………....6**

**2.1 DATA CLEANING AND PREPROCESSING………………………………….....6**

1. **METHOD OF ANALYSIS……………………………………………………………....7**
   1. **EDA………………………………………………………………………………......7**
   2. **MACHINE LEARNING MODEL…………………………………………….…...9**
   3. **DASHBOARD……………………………………………………………………...10**
2. **RESULTS………...………………………………………………………………….......12**
3. **CHALLENGES…….………………………………………………...…………………12**
4. **CONCLUSION…………………………………………………………………………13**
5. **LINK FOR GITHUB…………………………………………………….......................13**
6. **REFERENCES……………………………………………………….............................13**

**List of Figures**

* + 1. **No. of. Machines………………..……………………………………………………….. 9**

**3.1.2 Scrap types for both Presses…………………………………………………………….. 9**

**3.1.3 Up and Downtime for both Presses…………………………………………………….. 9**

**3.1.4 Press 1 Downtime Reasons……………………………………………………………. 10**

**3.1.5 Press 2 Downtime Reasons……………………………………………………………. 10**

**3.2.1 Machine Learning Model(snipped code)……………………………………………... 11**

**3.3.1 Press 1 Dashboard……...……………………………………………………………… 12**

**3.3.2 Press 2 Dashboard……...……………………………………………………………… 12**

1. **Introduction**

Machine downtime is a serious problem that affects companies of all sizes and sectors. Downtime is when a machine or piece of equipment is not in use, whether it's because of planned maintenance, emergency repairs, or unforeseen problems. This may result in wasted production time, missed deadlines, lower-quality products, and higher expenses. Therefore, companies must comprehend the causes and implications of machine downtime and create practical plans to lessen its effects.

A new generation of solutions to the problem of machine downtime has emerged in recent years as a result of the considerable technological breakthroughs brought about by Industry. Predictive maintenance programmes, for instance, utilize machine learning algorithms to forecast when a piece of equipment is likely to break down, allowing companies to take preventive action and reduce downtime.

Despite these developments, machine downtime continues to be a major problem for organizations, especially in sectors like manufacturing where uptime is essential for achieving production goals and guaranteeing product quality. To minimize the effects of downtime, organizations must regularly monitor and analyze the operation of their equipment, identify the underlying reasons, and put into practice efficient procedures.

To reduce the negative consequences of machine downtime on organizations, this research strives to shed light on its causes and effects. The study will entail examining market reports, case studies, and original research data to pinpoint the major reasons for downtime and assess the most recent technological developments, such as preventative maintenance approaches. The project's results will be helpful to companies trying to increase equipment performance and decrease downtime, as well as to academics and industry experts with an interest in industrial maintenance and dependability.

In the parts that follow, we will discuss our study of the factors that contribute to machine downtime, their effects, suggestions for tactics to reduce downtime, and a review of recent technological developments.

1. **Data Description**

To represent factories, equipment, and specific data points recorded by sensors, ShiftWorx uses a tree structure called Data Sources. The type\_id values, which range from 1 to 8, are used by the system to categorize the different types of data sources. Users starting jobs, status codes, or operators clocking in/out all result in the creation of data source attachments. The attachments serve as a record of a person's time spent working at the Data Source and include dates indicating when each attachment began and concluded. A list of the factory's operators may be found in the system's Users.csv file, and each user has a matching user\_id in the Users Attachments.csv file.

Status codes are employed to contextualize a machine's operational condition. Each of these codes has a parent\_id that points to its parent Status Code and is organized in a tree style. Each company uses the same two root codes for the system, Uptime and Downtime. Status codes are mostly used to describe downtime; uptime codes are rarely utilized. The Jobs.csv file includes a list of every job offered by the business as well as any job attachments. Data Source attachments in the CSV file show when tasks were started and ended on the target Data Source. A Jobs.csv ID is mapped to the job\_id column.

The system has a Scrap Types.csv file that contains a list of scrap types defined by the user. The Scrap Events.csv file contains data about scrap made, and the scrap\_type\_id column maps to an ID from Scrap Types.csv. Scrap should not be affected by a Job's multiplier. The system has a State Mask that tells users what state the machine was in before the current state.

**2.1 Data Cleaning and Preprocessing:**

To prepare our data for analysis, we conducted a thorough data cleaning and preprocessing process. The initial state of the data was unstructured and spread across 12 separate files, making it difficult to work with. To address this issue, we implemented ETL (extract, transform, load) techniques to normalize the data files and create a more structured format.

Our data preprocessing steps involved several key activities, including data quality assessment and cleansing, data integration and transformation, and data enrichment where applicable. We also ensured that our data was consistent and accurate by identifying and correcting any errors or inconsistencies that were present.

Through our ETL process, we were able to unify our data into a single, structured format that was optimized for analysis. This allowed us to better understand our data and extract meaningful insights that could be used to support our project objectives.

As a result of our data preprocessing and normalization efforts, we were able to consolidate our initial 12 separate files into six main files for use in our project. The following is a list of file names that are used further.

1. Company and Machine Information
2. Data Source Attachments
3. Users
4. Status Codes
5. Jobs
6. Scrap
7. **Methods of Analysis**

This section describes the various techniques and tools that were used to analyze and interpret data related to our project which are listed below:

* EDA (Exploratory Data Analysis)
* Machine Learning Model
* Dashboard in PowerBI

**3.1 EDA (Exploratory Data Analysis)**

In our project, we used Exploratory Data Analysis (EDA) as a method of analysis to identify the factors and patterns that helped us monitor machine downtime. EDA is a process of analyzing and summarizing data sets to gain insights into the data and its underlying structure. Through EDA, we were able to identify key factors that contributed to machine downtime. To visualize and communicate our findings, we created several data visualizations using Tableau.

Firstly, we had data from two machines that were being monitored. These machines were referred to as Press 1 and Press 2.

Chart

Description automatically generated

*Figure 3.1.1: No. of machines*

Next, we have identified the types of scrap generated from each press and the maximum type of scrap generated.

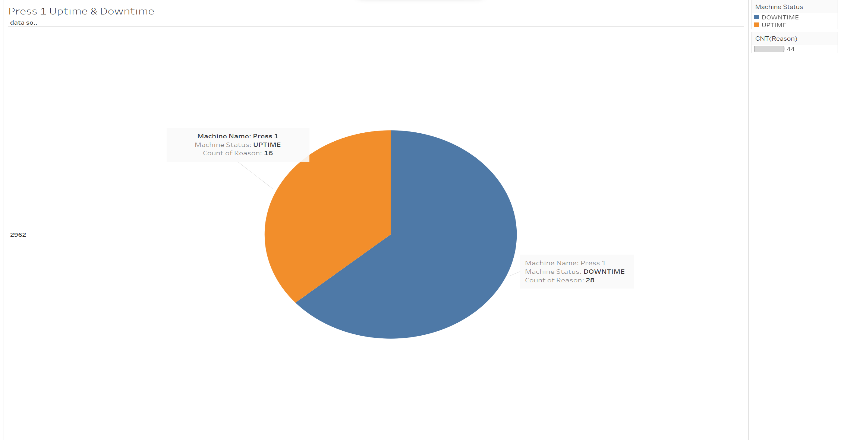
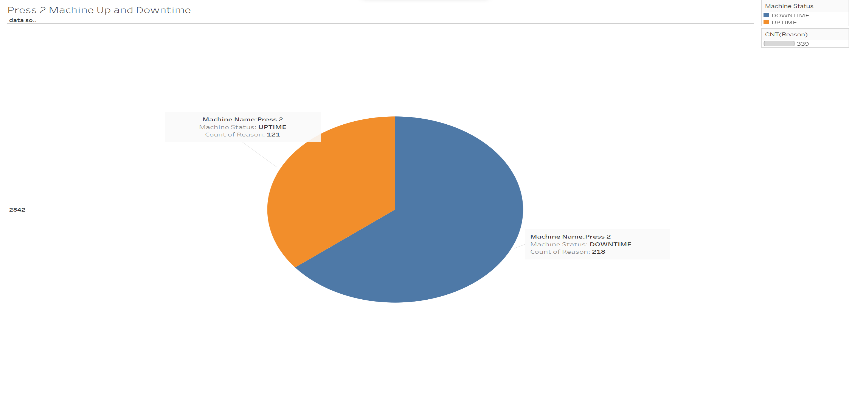
Chart, bar chart

Description automatically generated

*Figure 3.1.2: Scrap types for both presses*

Furthermore, a pie chart was used to illustrate the total up and downtime for Press 1 and Press 2, with Press 1 having 16 uptime and 28 downtimes and Press 2 having 121 uptime and 218 downtimes.

Press 1: Press 2:

*Figure 3.1.3: Up and Downtime for both presses*

Next, the analysis showed that for Press 1, the highest contributing reason for machine downtime was startup (technical waiting), among various reasons for downtime.

Chart, bar chart

Description automatically generated

*Figure 3.1.4: Press 1 Downtime Reasons*

For Press 2, the analysis revealed that Processing Tech (problems that occurred in the middle of operation) was the main reason for machine downtime, compared to other reasons for downtime.

Chart

Description automatically generated

*Figure 3.1.5: Press 2 Downtime Reasons*

**3.2 Machine Learning Model**

In our project, we developed a prototype model to address our problem. Even though we were unable to achieve a high degree of accuracy and minimize the loss in our initial attempts, we were able to create a functional prototype model based on the data that we had available. We plan to continue refining and improving the model as we obtain additional data in the future. The prototype serves as a foundation for further development and optimization, and we are optimistic that with sufficient data and tuning, we will be able to achieve our desired performance levels.

For our project, we opted to use the logistic regression machine learning method. This approach is commonly employed in binary classification problems where the aim is to predict a binary output variable based on one or more input variables. Our objective is to predict whether a machine is operating normally or not, which also constitutes a binary classification problem. As such, we determined that logistic regression is a fitting method for our needs. Notably, logistic regression is a straightforward and explainable model that is frequently utilized as a baseline for classification tasks.

Text

Description automatically generated

*Figure 3.2.1: Machine Learning Model(snippet code)*

Above snippet code is used to train and evaluate a logistic regression model on a dataset that contains information about machine statuses, such as the time a machine was attached and detached, the duration of the attachment, the reason for attachment, and the machine status (either "DOWNTIME" or "UP").

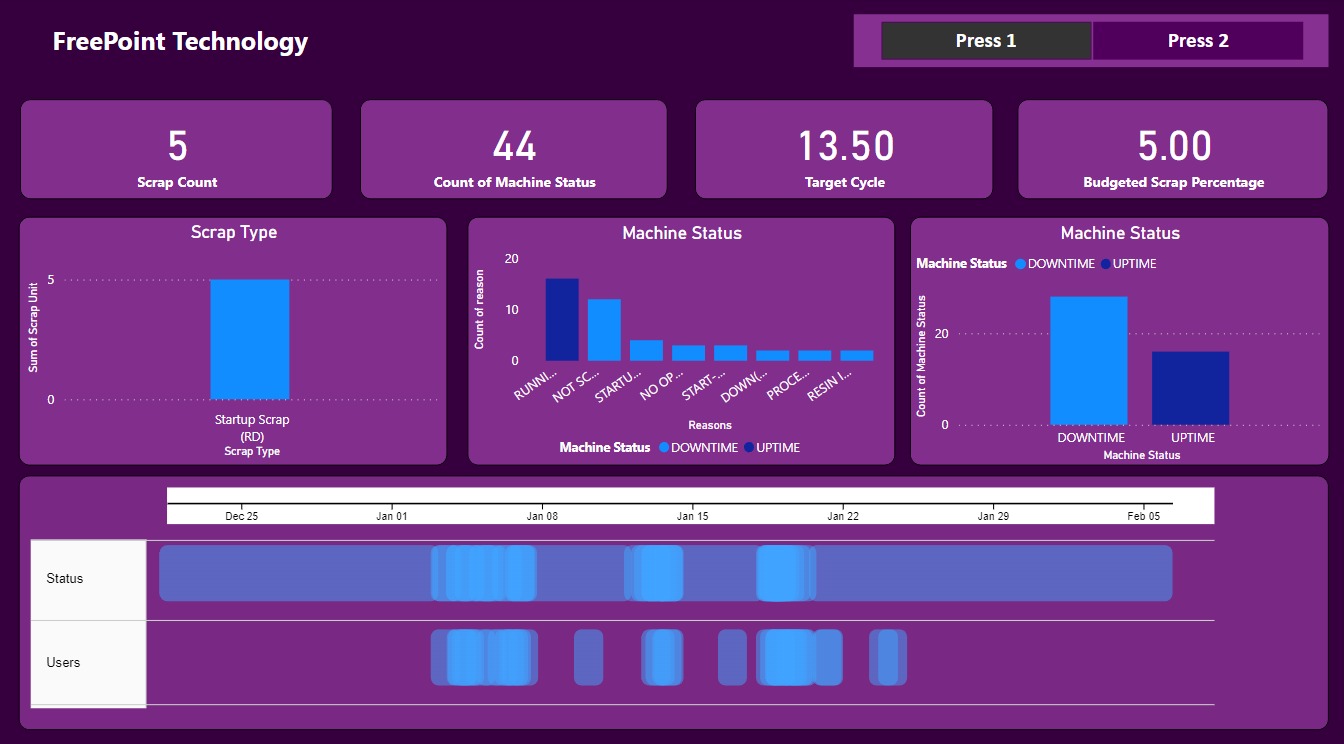
The code first preprocesses the data by converting date columns to DateTime format, calculating the duration of each attachment, converting categorical variables to numerical using label encoding, and dropping rows with missing values. The data is then split into training and testing sets.

A logistic regression model is then trained on the training set, and the model is evaluated on the testing set using accuracy, precision, recall, and F1 score metrics.

**3.3 Dashboard Using PowerBI**

The dashboard for our project provides an overview of Press 1 and Press 2, with a focus on machine status and scrap. It allows us to switch between Press 1 and Press 2 using a filter. This feature enhances the usability of the dashboard, as it enables us to view machine-specific metrics and performance data for each press. The dashboard aims to provide a quick and easy way for users to monitor the performance of both presses.

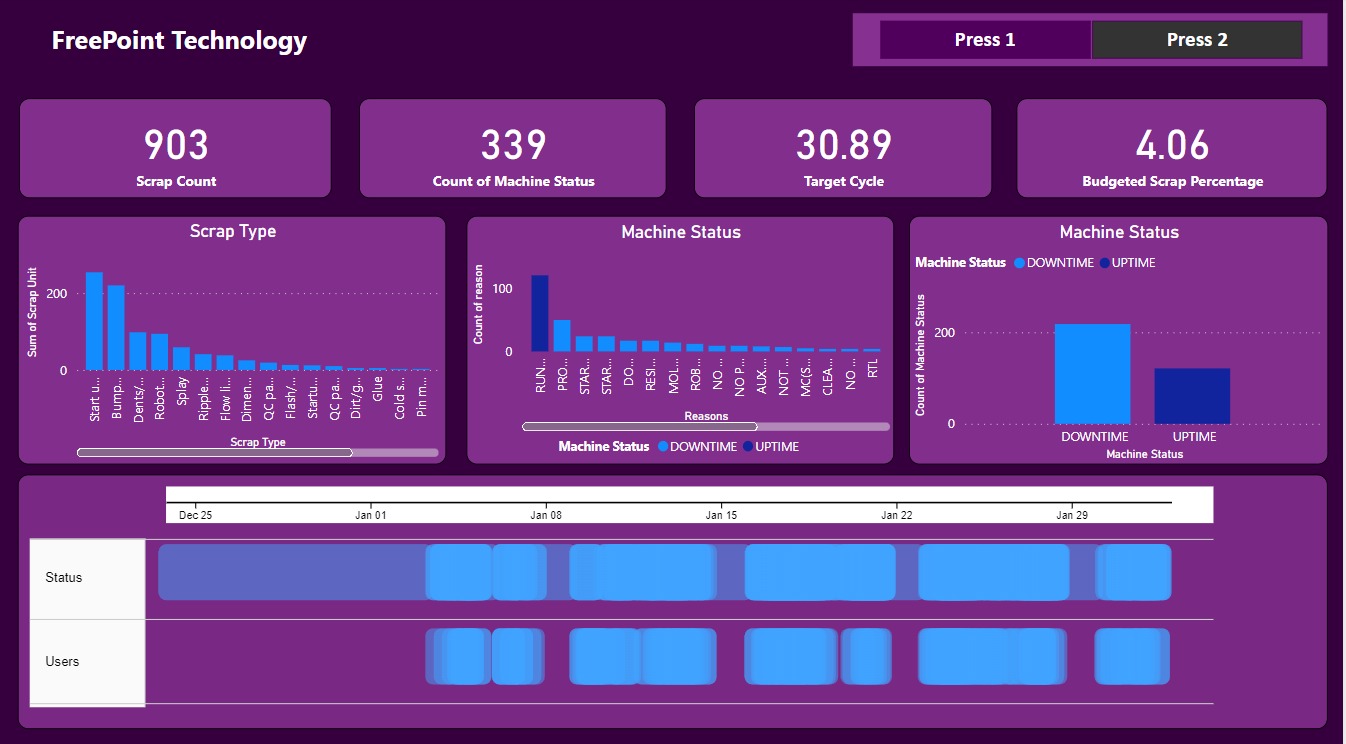
**Dashboard Press1:**



*Figure 3.3.1: Press 1 Dashboard*

Dashboard 1 provides an overview of Press 1, displaying machine status and scrap data. It also displays machine status information when the user is assigned to it. This ensures that when the user is assigned to the machine the status is usually UP. The dashboard is designed to keep users in sync with machine status, enabling them to monitor performance and make informed decisions promptly.

**Dashboard Press2:**



*Figure 3.3.2: Press 2 Dashboard*

Moreover, the dashboard displays the machine's status when a user is assigned to it, with the expectation that the machine will predominantly remain in the UP state. The primary objective of this dashboard is to keep users up-to-date with the machine's performance, facilitating their ability to track progress and make informed decisions promptly.

1. **Result**

In the results of our study, we found that the machine learning model we developed had an average evaluation matrix and accuracy is 0.25%, which is likely due to it being a prototype model. We believe that if we have access to more data, we could further tune the model to improve its performance.

To help visualize our findings, we have created a dashboard which provides an overview of Press 1 and Press 2. It also includes the maximum tolerance time of 13.60 minutes for Press 1 and a budgeted scrap rate of 5%. Dashboard for press 2 offers a summary of Press 2 by presenting its machine status and scrap data, along with a maximum tolerance time of 30.89 minutes and a budgeted scrap rate of 4.06%.

1. **Challenges**

The current dataset may not be sufficient to accurately train a machine-learning model for this problem. This is because the data appears to be biased towards a particular label (i.e., label 2) and may not be representative of the true distribution of data. Additionally, the dataset only spans a single month, which may not capture all the variations and patterns that could be present in the data over a longer period.

Furthermore, the current dataset only provides information on the attachment and detachment times of the sensors, and the reason for the status change. This limited information may not be enough to capture all the relevant features and factors that could influence the status of the sensors. For example, there may be other environmental factors, sensor readings, or operational conditions that could impact the sensor status but are not captured in the dataset.

In summary, while the current dataset provides some information about the sensor status, it may not be sufficient for accurate machine learning model training. To build a more robust and accurate model, a larger and more diverse dataset, with continuous sensor data and additional relevant features, would be needed.

1. **Conclusion**

Based on our observations during the study, we have found that both Press 1 and Press 2 have experienced more downtime than uptime during the one month. This highlights the need to focus on reducing downtime for both machines.

In particular, we have identified that Press 2 produces a higher volume of scrap, including various types of scrap, which requires more attention in terms of reduction efforts. In contrast, press 1 produces a single type of scrap, making it easier to address effectively.

Assigning operators to the machines can also help to reduce downtime, as we have found that the machines are less likely to experience downtime when they are being attended to by an operator. Additionally, we have established maximum tolerance downtime values for both machines to ensure optimal efficiency and prevent downtime from exceeding a certain threshold.

Overall, to improve the efficiency and productivity of the machines, it is necessary to focus on reducing downtime for both Press 1 and Press 2. By minimizing scrap production, assigning operators to machines as needed, and adhering to the maximum tolerance downtime values, we can help to ensure that downtime is kept to a minimum and that the machines are operating at their full potential.

1. **Link for GitHub Repository**

[**https://github.com/Mirenpatel1999/Capstone-Project**](https://github.com/Mirenpatel1999/Capstone-Project)

1. **References**

* [**https://github.com/mapr-demos/predictive-maintenance/blob/master/notebooks/jupyter/LSTM%20For%20Predictive%20Maintenance-ian01.ipynb**](https://github.com/mapr-demos/predictive-maintenance/blob/master/notebooks/jupyter/LSTM%20For%20Predictive%20Maintenance-ian01.ipynb)
* [**https://rjunaidraza.medium.com/predictive-maintenance-using-machine-learning-lstm-python-373396c6bb82**](https://rjunaidraza.medium.com/predictive-maintenance-using-machine-learning-lstm-python-373396c6bb82)
* [**https://towardsdatascience.com/lstm-for-predictive-maintenance-on-pump-sensor-data-b43486eb3210**](https://towardsdatascience.com/lstm-for-predictive-maintenance-on-pump-sensor-data-b43486eb3210)
* [**https://www.engineering.com/story/how-to-reduce-machining-time-with-3-axis-turning**](https://www.engineering.com/story/how-to-reduce-machining-time-with-3-axis-turning)
* [**https://mechmatrix.in/how-to-reduce-cycle-time-without-reducing-tool-life-in-cnc-machining/**](https://mechmatrix.in/how-to-reduce-cycle-time-without-reducing-tool-life-in-cnc-machining/)
* [**https://www.meadenmoore.com/blog/consulting/benefits-of-reducing-machine-setup-time#:~:text=Although%20decreasing%20run%20times%20reduces,flexibility%20to%20meet%20customer%20demands**](https://www.meadenmoore.com/blog/consulting/benefits-of-reducing-machine-setup-time)**.**
* [**https://www.machinemetrics.com/blog/reduce-downtime-manufacturing**](https://www.machinemetrics.com/blog/reduce-downtime-manufacturing)