**DAB 402 – CAPSTONE PROJECT**

**DATA ASSESSMENT-2**

**PROJECT TOPIC:**

**Reducing Machine Down Time**

**Group**:- **7**

**Group Members:**

**Mirenkumar Patel (**0782036)

**Zeelkumar Mandpara** (0790824)

**Nirmit pareliya**(0785084)

**Miloni Patel**(0788500)

**Harsh Patidar**(0791820)

**Reducing Machine Downtime:**

**Description:**

Reducing machine downtime is the process of cutting down on the amount of time a machine is not in use. Downtime may be brought on by a number of things, including mechanical issues, operator mistakes, and the need for repair, and it can cost firms a lot in terms of lost output, lower productivity, and higher maintenance costs.

Businesses may use a number of tactics to lower machine downtime, including developing predictive maintenance programmes, enhancing operator training and monitoring, and utilizing machine learning algorithms to anticipate and avoid downtime situations.

In order to detect trends and anomalies that can point to potential breakdowns or the need for maintenance, machine learning algorithms can evaluate previous data on the functioning of a machine. This can help firms reduce downtime by scheduling maintenance within pre-planned downtime windows or anticipating and taking care of maintenance needs before a machine breaks down.

Businesses may boost production productivity, save maintenance costs, and enhance overall equipment performance by decreasing machine downtime. By providing high-quality goods and services, may lead to considerable cost reductions, better profitability, and enhanced customer happiness.

**Data Assessment Steps:**

* . **Data Source(s) and Gathering:** The collecting and use of data are crucial steps in the data assessment process. Here are some steps involved in identifying data sources and gathering data:
  + **Determine possible sources of data:** Identify the data sources that are pertinent to the study topic or business challenge. These sources consist of external sources.
  + **Get authorization and access:** Request authorization and access to the information sources. This might entail negotiating contracts, signing contracts, or getting permission.
  + **Gather the data:** Get the information from the suggested sources. This might entail data collecting, data scraping, or data extraction. In our case, it was data collecting.
  + **Validate the data:** Check the precision and thoroughness of the information gathered. Checking for missing data, outliers, or contradictions may be necessary for this.
  + **Data preparation**: Data preparation is the process of preparing data for analysis. This might entail data normalization, data transformation, or data cleansing.
  + **Integrate the data:** Combine the data gathered with information from other reliable sources. This can entail data connecting or data fusion.
  + **Data storage:** Data should be kept in a safe, well-organized location that makes it simple to access and retrieve. Using a database or data warehouse may be required for this.
  + **Record the data:** The data sources, data gathering techniques, and data preparation stages should all be documented. This makes the data more transparent and replicable.

These were some of the steps required to do Data Gathering. But as we are having external sponsors most of the work was done by our sponsors in order to maintain their confidentiality.

* **Dataset Features/Attributes:** The reducing machine downtime dataset may include a number of features/attributes that describe the downtime incidents, including the machine ID, machine type, date and time of the downtime, how long it lasted, what caused it to happen, where the machine was located, and the technician's name.
* **Dataset Quality:** The correctness, comprehensiveness, and consistency of the data are referred to as the dataset's quality. The dataset for decreasing machine downtime must be devoid of mistakes, duplication, and inconsistencies, and it properly reflects the instances of downtime. To guarantee the quality and comprehensiveness of the data, the dataset is also updated on a regular basis.
* **Dataset Fitness:** The dataset's suitability for the goal of lowering machine downtime is referred to as its fitness. The dataset is having pertinent traits or properties that may be used to pinpoint the source of downtime issues, and it must is sufficiently large to hold enough data for analysis.
* **Dataset Usability:** How readily a dataset can be accessed, comprehended, and used is referred to as its usability. The dataset is well-structured, well-documented, and in an accessible manner. Furthermore included with the dataset are detailed instructions on how to use it.
* **Missing Values:** Missing values are instances of missing or insufficient data in the dataset. Missing numbers in the minimizing machine downtime dataset can happen for a number of reasons, including insufficient data or data input mistakes. To guarantee the correctness and dependability of the analysis, it is crucial to handle missing values effectively. Imputation, deletion, and modelling are often used techniques for dealing with missing variables.
* **Dataset Terms of Use:**

In order to manage datasets used in a responsible and ethical manner, dataset terms of use are essential. These terms outline the terms and conditions for accessing, using and disseminating the dataset by defining the rights and responsibilities of users and data suppliers. The approved uses of the data, including restrictions on who may use the data, the intended use for the data, and whether the data may be used for commercial or non-commercial reasons, are outlined in the thorough dataset terms of use agreement. The agreement also outlines who owns the data, the intellectual property rights attached to it, and any restrictions on its modification, copying, or dissemination.

The conditions to use are to forbid certain actions like sharing data with unauthorized parties or using it for illicit purposes in order to ensure correct usage of the information. Any attribution specifications, such as acknowledging the data source in academic papers or other publications, are also included in the agreement. A disclaimer of warranties or indemnity clauses that restrict the data provider or owner's obligation for errors or inaccuracies in the data is included in the terms of usage to reduce liability.

In conclusion, dataset terms of use are an important aspect of using datasets. They establish a precise framework for the permitted use of the data, place restrictions on responsibility, and encourage moral and responsible usage. To guarantee the correct use of the dataset and prevent any legal difficulties or breaches, researchers and analysts must carefully study and adhere to the conditions of use.

* **Dataset Reproducibility:**

The dataset is to be well curated and recorded in order to achieve dataset repeatability in the context of minimizing machine downtime. This entails recording any data cleansing or pre-treatment activities as well as the data analysis approach. The software and hardware that are utilized for the study, as well as any pertinent settings or parameters, should all be documented.

It is advised to log changes to the dataset and analytic code using version control tools to ensure repeatability. This will make modifications made throughout the analysis process easier to reproduce and well-document. By enabling others to confirm the findings and repeat the investigation, the usage of code sharing may also support repeatability.

Lastly, it is crucial to guarantee the dataset's quality, including the data's completeness and correctness. This can be accomplished by carefully cleaning and preparing the data, taking care of any missing or inconsistent data, and all of the above. Reproducibility may be attained through supporting high-quality, thoroughly documented datasets and analysis, fostering openness and accountability, and decreasing machine downtime.

* **Dataset Storage & Handling:**

None of our datasets exceeds 1MB and as such we don’t have to make special provisions for the datasets storage as it’s small in size and transferable. If there is any additional requirement for the data then we are assured that we will be provided with the API through which we can access the data which we need.

**REFERENCES:**

<https://www.techtarget.com/searchdatamanagement/definition/data-quality>

<https://pair.withgoogle.com/chapter/data-collection/>

<https://www.altexsoft.com/blog/datascience/preparing-your-dataset-for-machine-learning-8-basic-techniques-that-make-your-data-better/>

<https://evocon.com/articles/downtime-tracking-and-analysis/>

<https://www.yokogawa.com/ca/library/resources/media-publications/downtime-analysis/>

<https://www.machinemetrics.com/blog/machine-downtime-analysis>

<https://labelyourdata.com/articles/predictive-maintenance-datasets>

<https://kb.ei3.com/machine-production-states/>