Predictive Maintenance for Elevators: Harnessing Sensor Data for Failure Prevention

Assignment 3

A.A Namith Nimlaka (31073484)

Project Description

All modern high-rise buildings highly rely on the use of elevators for efficient transportation between their floors. However, there seem to be many points of failures in these machines which pose a great safety risk to occupants even to this day.
The two biggest points of failure in an elevator as stated by Sarkar et al. (2022), are the failure of the elevator door component and the Operational Runtime failure.
According to Yu et al. (2020), **Elevator Door failure** has been the most common point of failure, from wasting time to severe claustrophobia this may pose major safety concerns to occupants.
Operational failure even though rarer can be even worse, this may occur due to unmaintained machinery which is prone to going unnoticed when wearing down as these components are hidden within the elevator and are only maintained at chosen intervals and not when required.

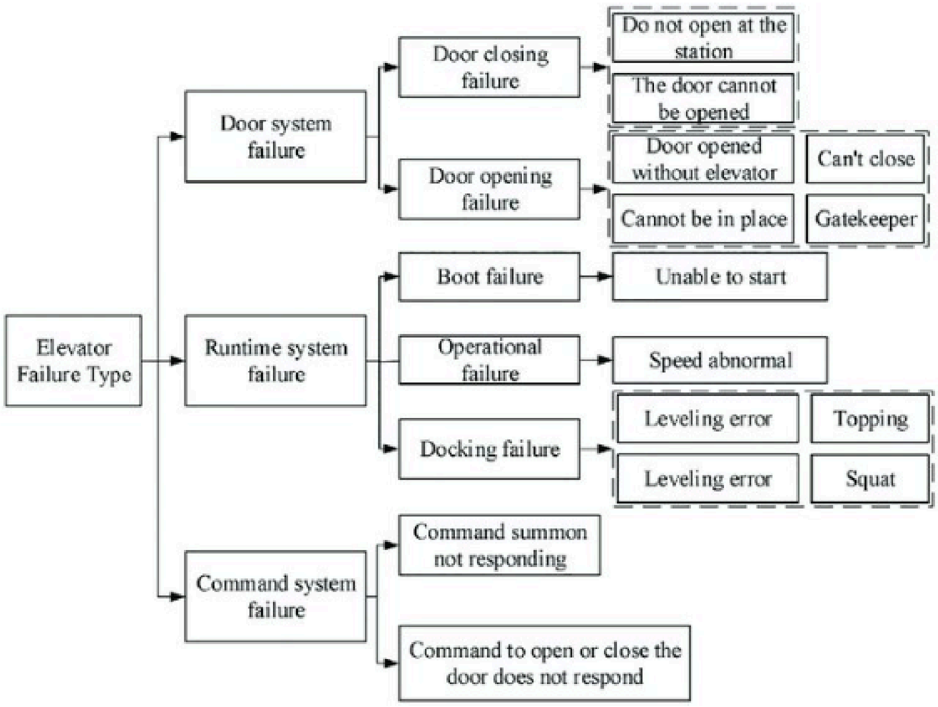


Figure 1: Major Elevator Failures by Yu et al. (2020).

Objective

This project focuses on the prevention of failures ever occurring, as they say, "The best approach to fix any issues is to prevent it from ever occurring". By using datasets to predict the maintenance of the components using their sensor values, the multiple different sensor readings for different instances of the elevator can be used to deduce the **RUL(Remaining Useful Life) of the components** and the **vibration of the elevator to retain the standards for establishmen**ts and allow proactive maintenance interventions whenever they may be required.

Currently, maintenance is performed in intervals from monthly to yearly, this has significant issues in either unnecessary downtime and costs or high rates of failure due to ignorance of maintenance, this application can prevent this by proactively using sensor readings to predict if the elevator would require maintenance greatly reducing downtime, costs and risks for potential users.

Data Science Roles and Responsibilities

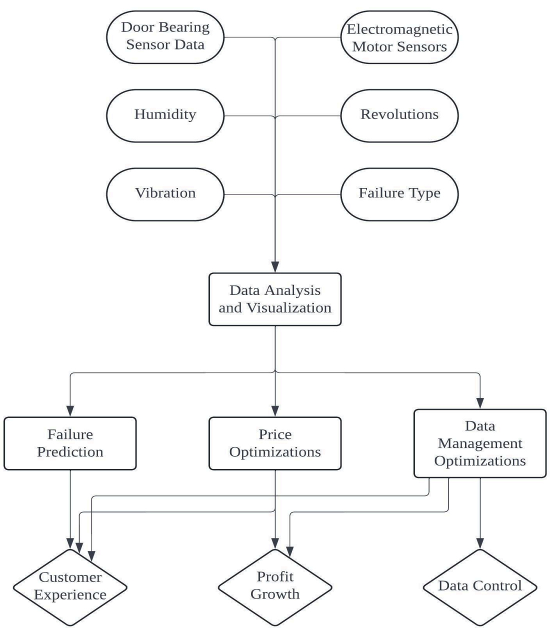
A **Data Engineer** would be required to automate the fetching of newer data from sensor readings and update the database as required to provide more accurate predictions with more quality data.

A **Data Analyst** would be required to visualise all the data from our database, they would need to wrangle this data transform it into a usable format and present it as a report that can display the findings, quality of the data and accuracy of the predictions to potential buyers.

A **System Architect** with development experience is required to integrate all the sensor data with the predictive maintenance algorithms to create the final application and UI.

A **Data Scientist** would analyze the sensor data and find the patterns and insights that would be required to build all the outliers for the failure of the elevator components.

Business Model



The project would mainly target the elevator manufacturers and elevator maintenance industry, any high-rise building which would have an elevator that is regularly maintained would see this as an optimal solution, offering predictive maintenance for the major elevator components such as the doors and bearings would provide the seekers minimal disruptions and required maintenance. It would be sold as **Software as a Service (SaaS)** with an application built using the same sensor data for predictive maintenance but will be then connected to each customer's system and their elevator sensors, the payments would be made **subscription-based**.

Figure 2: Influence Diagram for a business model.

Benefits

1. Reduced Downtime: Proactive maintenance minimizes unplanned stops and elevator downtime, ensuring smooth operations during peak usage periods.
2. Cost Savings: By predicting failures in advance, building owners can avoid costly emergency repairs and extend the lifespan of elevator systems, resulting in significant cost savings over time.
3. Enhanced Safety: Predictive maintenance improves elevator safety by identifying potential issues before they escalate into critical failures, reducing the risk of accidents and injuries.
4. Customer Satisfaction: Reliable elevator performance enhances the overall tenant experience, leading to higher satisfaction rates and improved tenant/customer retention.

Challenges

1. Data Integration: Integrating data from diverse IoT sensors and ensuring data quality and consistency across all types of elevators may pose technical challenges.
2. Model Accuracy: Developing accurate predictive maintenance models requires robust feature engineering and validation against a very large set of real-world failure data.
3. Deployment and Scalability: Implementing predictive maintenance solutions across large-scale elevator networks requires scalable deployment strategies and ongoing monitoring for system optimization.

Characterising and Analysing data

Data Sources

Elevator Door Failure Dataset
The dataset focusing on elevator door failures is crucial for predicting and preventing door-related malfunctions in elevators. This dataset can be sourced primarily from elevator manufacturers, maintenance companies or open data. These entities typically maintain comprehensive records of elevator failures, including sensor readings from ball bearings and other components during stress testing or reported incidents. By analyzing these sensor readings, we can identify patterns and indicators of impending door failures, allowing for proactive maintenance measures to be implemented.

Vibration Prediction Dataset
The vibration prediction dataset is essential for estimating the absolute value of vibration in elevator bearings. This dataset can be obtained from manufacturers or open data specializing in elevator systems. Absolute vibration measurements involve using eddy current sensors for relative shaft vibration and accelerometers or velocity sensors for absolute bearing housing vibration. By subtracting the relative vibration from the correct phase of the absolute vibration, we can derive the absolute shaft vibration. This dataset aids in setting vibration prediction standards tailored to different establishments, ensuring that maintenance activities are scheduled proactively based on real-time vibration data.

Characteristics of Data

The datasets exemplify the four V's of big data. The **volume** of data is substantial, reflecting the extensive sampling inherent in sensor data collection across a multitude of elevator systems. The **variety** within the datasets is significant, showing the diverse sensor configurations utilized in different elevator systems. The **velocity** is applied through real-time data capture during peak elevator usage periods and ensures that the application remains responsive to elevator usage patterns. **Veracity**, integrity and accuracy of the data are crucial, given the potential for noise in sensor data, cleaning and preprocessing are essential to maintain data quality, ensuring the reliability of the predictive maintenance models.

Data Processing and Storage

In data processing, we first clean the data by **removing all NA, empty, or invalid data**, ensuring that only correct ranges are retained to eliminate corrupted or mistaken values. For the door failure dataset, we added a new column called "failure" to generalize failure types into two categories as a binary outcome, as the specific type of failure is not critical for maintenance decisions.

We then split both the datasets into **80% for training and 20% for testing**. During training, columns like UID, ID and failure type are excluded to avoid bias and remove columns unrelated to the target predictions.In terms of technology, data processing is primarily conducted using **R in Rstudio** along with analysis and visualisation packages for initial analysis, preprocessing, visualization, and statistical modelling. For storage, the initial utilization of **CSV files** offers simplicity and accessibility. However, plans are underway to transition to SQL databases like **PostgreSQL** to accommodate the increasing volume of data efficiently.

Statistical Methods and Data Analysis

Door Failure Dataset

Method
We will use **a classification tree** to predict door failures. This method is suitable because the door failure is a categorical outcome (either failure or no failure) based on sensor data. A classification tree can effectively handle this type of prediction by creating decision rules from the data features.

Analysis
To evaluate the performance of the classification tree, we will employ two different splitting criteria: Gini impurity and entropy. These criteria will be used to construct two separate models. After training the models, we will assess their performance using a **confusion table**, which will help us understand the accuracy and the ability of each model to correctly predict door failures on the test dataset.

Accuracy = Number of correct predictions / Total number of predictions

Vibration Dataset

Method
A **regression model** will be used for the vibration dataset because the goal is to predict a continuous vibration value based on sensor data. This approach is suitable for modelling and predicting numerical values.

Analysis
In addition to the regression model, two baseline models will be created to provide a benchmark for comparison. The first baseline model will predict the mean of the vibration values, and the second baseline model will predict the median. The performance of these models will be evaluated using the **mean squared error (MSE)** metric. The MSE of the regression model will then be compared to the MSEs of the baseline models to determine if there is a significant improvement.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Demonstration

Identified Dataset

Our project focuses on leveraging two main datasets to predict and prevent elevator failures. The dataset by Mr Wild (2023) for the sensor readings for predicting elevator door failure based on the available data of failure records collected from various sensor outputs and a second dataset by Bansal (2022) from the Huawei German Research Center for the operation data, in the form of time series sampled at 4Hz in high-peak and evening elevator usage in a building (between 16:30 and 23:30). This considers mainly Electromechanical sensors, Ambiance and targets to predict the absolute value of Vibration which can be used to estimate when maintenance for the bearings may be required.

Both datasets, the Door Failure Dataset and the Vibration Dataset adhere to the principles of the 4Vs of big data. They demonstrate significant volume with many observations of sensor data. Both datasets exhibit variety, as they encompass diverse sensor data. The datasets showcase velocity, with the potential for real-time data updates and continuous growth. And ensure veracity by accurately capturing and representing the relevant factors influencing elevator performance.

Dataset Analysis

Door Failure Dataset

```
# Check the structure of the failure dataset
glimpse(failure_raw)
```

```
## Rows: 20,009
## Columns: 10
## $ UID            <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, …
## $ cycle          <chr> "L", "M", "H", "L", "L", "L", "H", "L", "M", "L", "L", …
## $ p1             <dbl> 1551, 1408, 1498, 1433, 1408, 1425, 1558, 1527, 1667, 1…
## $ p2             <dbl> 42.8, 46.3, 49.4, 39.5, 40.0, 41.9, 42.4, 40.2, 28.6, 2…
## $ p3             <dbl> 0, 3, 5, 7, 9, 11, 14, 16, 18, 21, 24, 29, 34, 37, 40, …
## $ s1             <dbl> 298.1, 298.2, 298.1, 298.2, 298.2, 298.1, 298.1, 298.1,…
## $ s2             <dbl> 308.6, 308.7, 308.5, 308.6, 308.7, 308.6, 308.6, 308.6,…
## $ s3             <dbl> 5475.852, 5475.852, 5475.704, 5475.704, 5475.704, 5475.…
## $ `failure type` <chr> "No failure", "No failure", "No failure", "No failure",…
## $ failure        <chr> "no failure", "no failure", "no failure", "no failure",…
```

Figure 3: Column field data for the door failure dataset.

The dataset has the 6 different door ball bearing sensor data p1-p3 and s1-s4 and the failure type of the failure that has occurred, it consists of **20009 observations**.

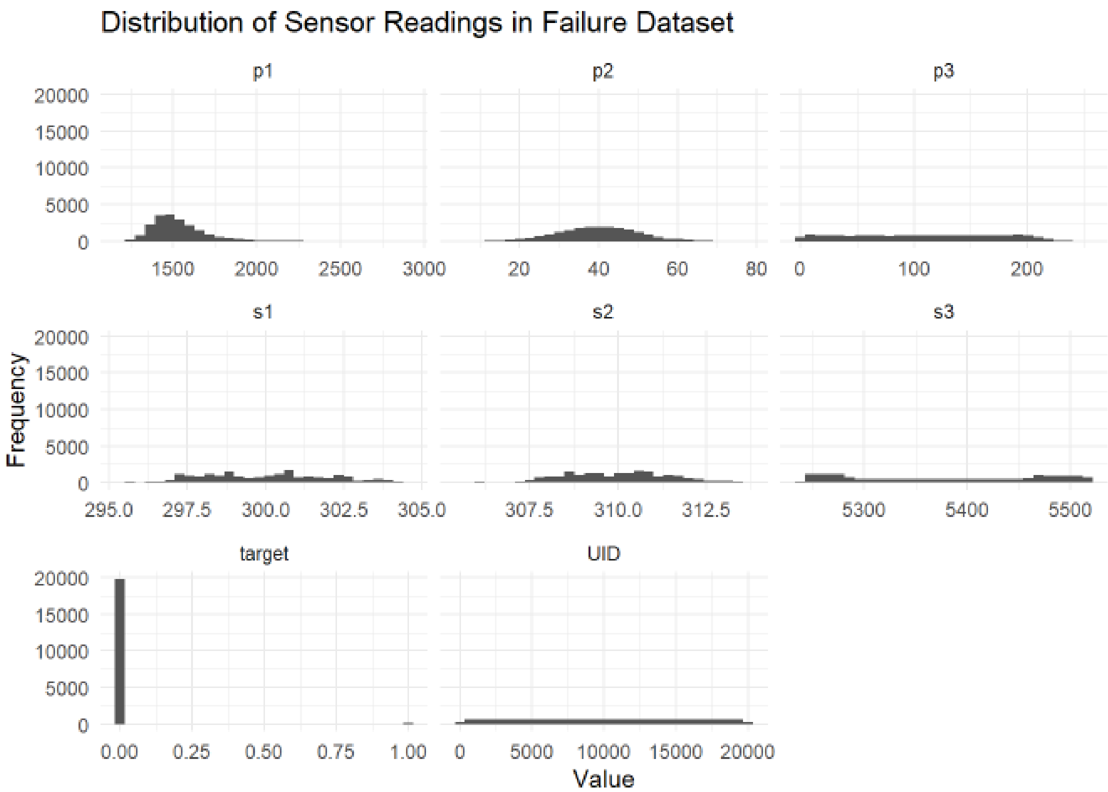


Figure 4: Distribution of sensor readings in Failure Dataset.

Vibration Dataset

```
# Check the structure of the vibration_rawset
glimpse(vibration_raw)
```

```
## Rows: 112,001
## Columns: 9
## $ ID          <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17,…
## $ revolutions <dbl> 93.744, 93.740, 93.736, 93.732, 93.729, 93.725, 93.721, 93…
## $ humidity    <dbl> 73.999, 73.999, 73.998, 73.998, 73.998, 73.997, 73.997, 73…
## $ vibration   <dbl> 18.00, 18.00, 18.00, 18.00, 18.00, 18.01, 18.01, 18.01, 18…
## $ x1          <dbl> 167.743, 167.739, 167.734, 167.730, 167.727, 167.722, 167.…
## $ x2          <dbl> 19.745, 19.741, 19.738, 19.734, 19.731, 19.728, 19.724, 19…
## $ x3          <dbl> 1.266828, 1.266774, 1.266737, 1.266683, 1.266642, 1.266605…
## $ x4          <dbl> 8787.938, 8787.188, 8786.438, 8785.688, 8785.125, 8784.376…
## $ x5          <dbl> 5475.852, 5475.852, 5475.704, 5475.704, 5475.704, 5475.556…
```

Figure 5: Column field data for the vibration dataset.

The dataset shows the 5 readings from the x1-x5 electromagnetic sensors, the humidity when the readings were taken and the revolutions the elevator has been through and consists of **112001 observations**.

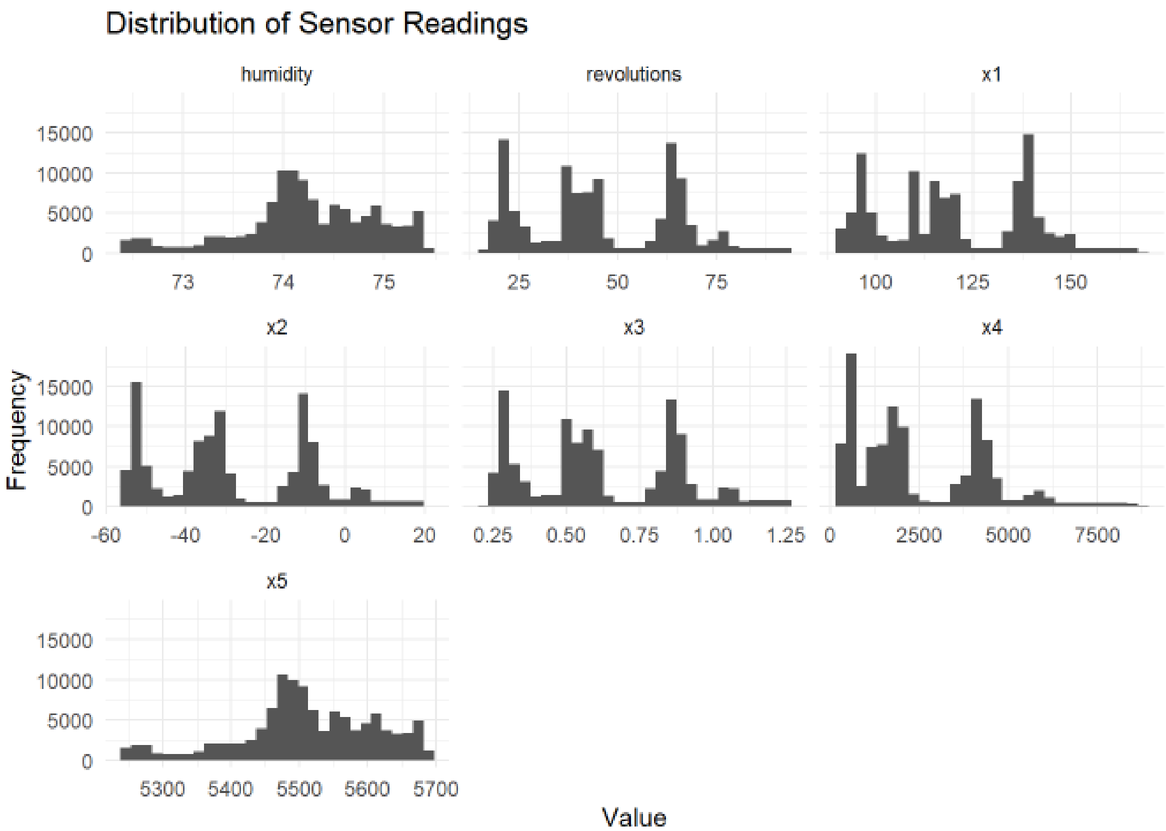


Figure 6: Distribution of sensor readings in Vibration Dataset.

Feasibility Analysis

Classification Tree Analysis

Gini

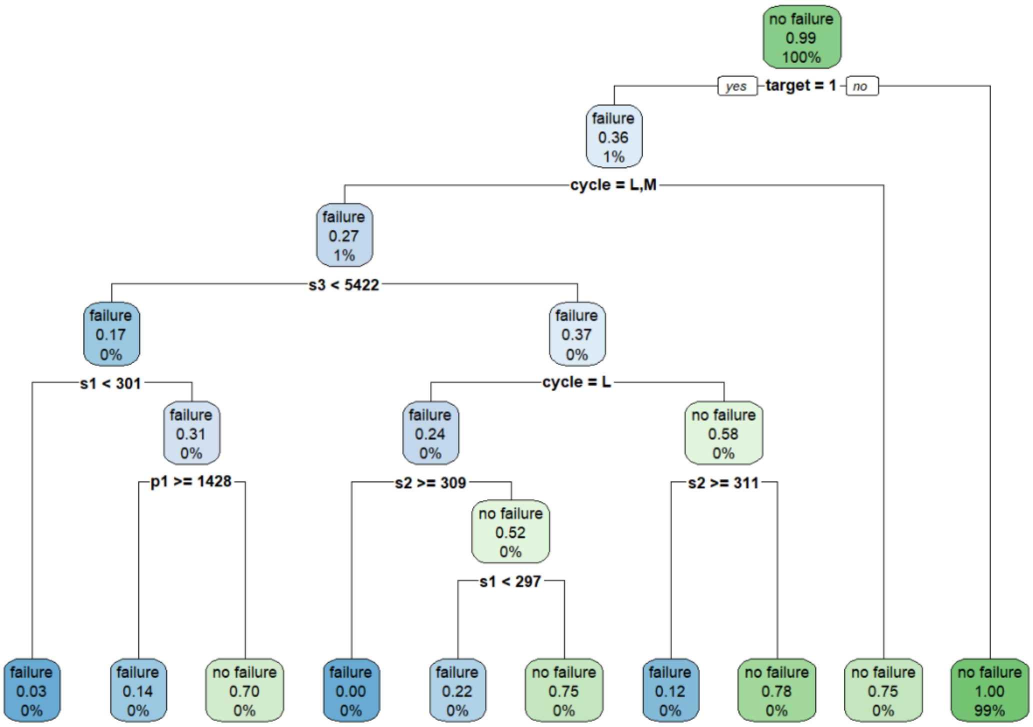


Figure 7: Classification Tree using Gini Split for Doof Failure Prediction.

```
##            pred_gini_test
##             failure no failure
##   failure         13          7
##   no failure       3       3979
```

Gini split has shown to have an **accuracy of 99.75%** in predicting the outcome from sensor readings but only **65% accurate failure prediction**.

Entropy

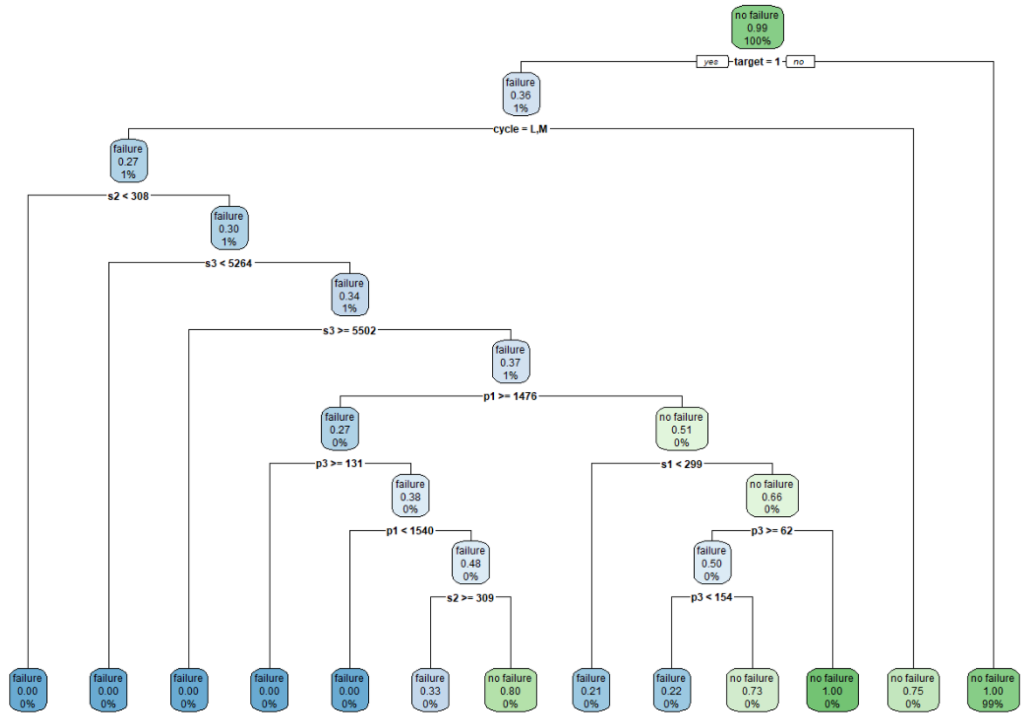


Figure 8: Classification Tree using Entropy Split for Doof Failure Prediction.

```
##              pred_entropy_test
##               failure no failure
##   failure          16          4
##   no failure        4       3978
```

Entropy split has shown to have an **accuracy of 99.8%** in outcome prediction and a higher **80% failure prediction** than that of the Gini split model and is therefore the selected method of training for the classification tree for failure prediction.

Regression Tree Analysis

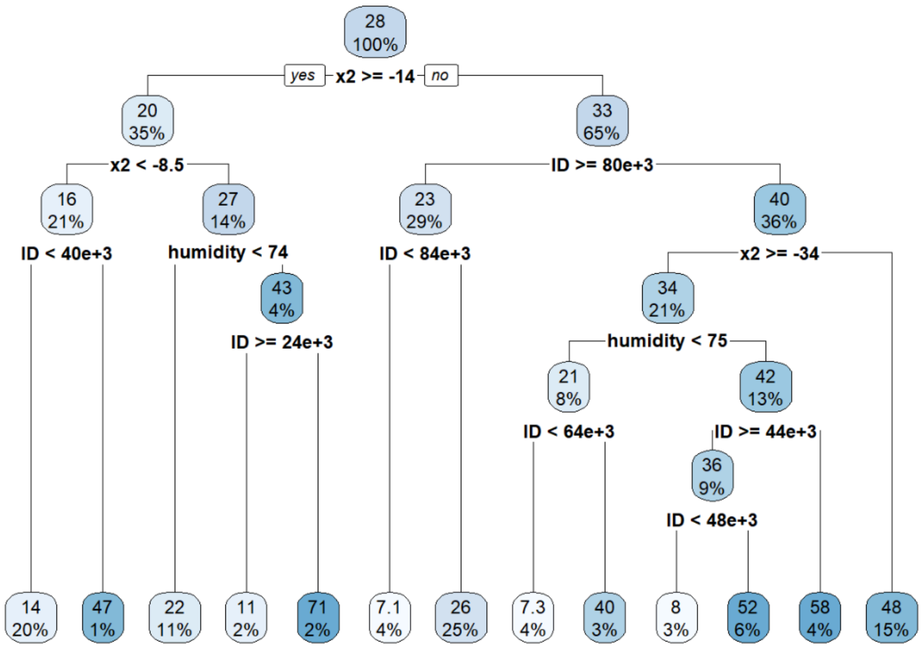


Figure 9: Regression Tree used for Absolute Vibration prediction.

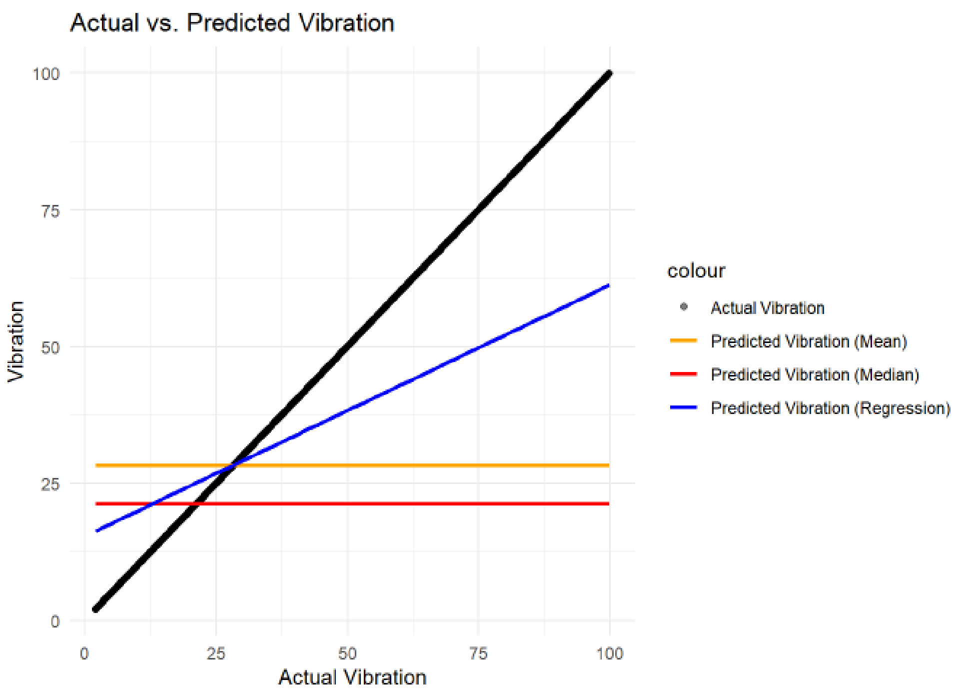


Figure 10: Predicted value comparison along with 2 baseline mean and median models.

The regression model has been tested alongside 2 other baseline models, the mean square error of the regression model is 319.28, while the mean model has an MSE of 594.39 and the median model has an MSE of 646.56, this shows that the regression model has a **significant improvement of 50% over the 2 baseline models** and does predict quite more accurate values to the actual values.

Standard, Data Governance and Management

Standard

Figure 11: CRISP-DM Standard steps.

The Chosen standard that we have followed is CRISP-DM (Cross-Industry Standard Process for Data Mining), including the steps of Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation of Model and finally Deployment. It follows an iterative approach, allowing for flexibility and iteration within each phase based on insights gained during the process.

Data Governance

As stated by Naguib et. al. (2024), we also focus on key aspects such as security, persistence, ethics, confidentiality, and regulatory compliance. Encryption is used to safeguard data, ensuring both its security and confidentiality. Regular backups are maintained to guarantee data persistence and recoverability. Additionally, we adhere to regulatory requirements, consistent terminology, and ethical standards, fostering a culture of responsibility within the organization. Each client receives all necessary data for their elevator models without bias, ensuring fair treatment and privacy. We strictly protect client data from any unauthorized access or leaks. Our commitment to ethics ensures that we handle all data responsibly and morally.

Data Management

Following the plan by Prokscha (2024), all retrieved data undergoes thorough preprocessing to eliminate empty or incorrect entries. We ensure that each column contains only unit data, avoiding any compound data. To maintain clarity and consistency, we follow strict file and field naming conventions, enabling an unambiguous understanding of the data. Additionally, we ensure that all variables are of appropriate types and that measurements use consistent units. Explicit and consistent definitions are provided for the failure type, ensuring clarity in data interpretation and analysis.

References

Sarkar, Sobhan & Vinay, Sammangi & Djeddi, Chawki & Maiti, Jhareswar. (2022). Classification and pattern extraction of incidents: A deep learning-based approach. Neural Computing and Applications. 10.1007/s00521-021-06780-3.

Yu, Jie & Hu, Bo. (2020). Influence of the combination of big data technology on the Spark platform with deep learning on elevator safety monitoring efficiency. PLOS ONE. 15. e0234824. 10.1371/journal.pone.0234824.

Prokscha, Susanne. (2024). Data Management Plans. 10.1201/9781003395621-3.

Naguib, Hend & Kassem, Hossam & Naem, Abd. (2024). The impact of IT governance and data governance on financial and non-financial performance. Future Business Journal. 10. 10.1186/s43093-024-00300-0.

Mr. Wild. (2023). Failure prediction of elevator component [Data set]. Kaggle. https://www.kaggle.com/datasets/mrwild/failure-prediction-of-elevator-component

Bansal, S. (2022). Elevator predictive maintenance dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/shivamb/elevator-predictive-maintenance-dataset