“TESTING INTELLIGENCE (TI)” AI PLATFORM FOR PREDICTIVE MAINTENANCE IN MEASUREMENT SYSTEMS FOR AUTOMOTIVE TESTING

Final Capstone Project

Oleg Moroka | MBAI21/MBA23 | 25.06.2023

Kyiv School of Economics

Abstract

Faculty of Business Education

Master’s of Business and Management in Artificial Intelligence and Data Analytics

“Testing Intelligence (TI)” AI platform for predictive maintenance in measurement systems for automotive testing

by Oleg Moroka

Predictive maintenance refers to using data analysis tools and techniques to identify anomalies in operations and potential equipment or process defects. By detecting these issues proactively, businesses can take corrective measures before they escalate into failures. Currently, conservative automotive industry either does not use any technique to predict system failure or uses statistical methods which either provide results too late orunreliables. To address these challenges, I propose to develop “Testing Intelligenca e (TI)” AI platform for predictive maintenance in measurement systems for automotive testing. TI will use state-of-the-art Artificial Intelligence (AI), which will be trained for specific businessdomainsn. Usage of the TI platform shall allow customers to reduce coststhe equipment availabilityment is increased but manual high-paid work is reduced. Also such solution frees up human and hardware resources which allows to be more time efficient in testing.

Acknowledgement

I would like to express my genuine gratitude to to my project supervisor, Professor Oleksandr Korobov, who believed in my idea and provided technical guidance.

Many thanks to the Honorary Academic Director of MBAI program, Professor .Oleksandr Romanko, who designed and brought top-level educational program into Ukraine.

And of course it would not be possible without all professors who shared their knowledge and experience with me during lectures.

KSE | Business Education

Table of Contents

Introduction

Automotive industry struggles for costs reduction. And predictive maintenance can help to determine the condition of equipment and predicting when maintenance should be performed. It can lead to major cost savings, higher predictability, and the increased availability of the systems. In predictive maintenance scenarios, data is collected over time to monitor the state of equipment. The goal is to find patterns that can help predict and ultimately prevent failures.

Some of the problems you can solve:

- Remaining useful life (RUL): RUL prediction gives you insights about when your machine will fail so you can schedule maintenance in advance
- Flagging irregular behavior: Anomaly detection through time series analysis
- Failure diagnosis and recommendation of mitigation or maintenance actions after failure

We will concentrate on anomaly detection which can provide immediate insights of either hardware or software failure. Anomaly detection involves identifying uncommon items, events, or observations that stand out due to their significant deviation from the majority of the data.

Automotive is quite conservative industry and currently mainly statistics methods are used for anomaly detection like Boundary/Limit Monitoring and Statistical Process Control (since 1924). See *Figure 1 Statistical methods for anomaly detection.*

However, those methods are not applicable to high-dynamic measurements as could provide false positive results. Those methods could provide false negative results when the feature is still within limits.

Those methods could detect anomaly **too late** when failure is already occurred. And as a result repair cost and lost production.

Those methods require manual effort of high-qualified human resources for root cause/insights analyze.

The proposed solution will be based on state-of-the-art Artificial Intelligence (AI).

In the context of industrial AI, the process referred to as "training" empowers ML algorithms to identify anomalies and explore correlations as they search for patterns within the diverse data feeds.

AI solution can provide the following benefit:

- Automation: analyze datasets, dynamically fine-tune normal behavior parameters and identify breaches in in patterns
- Real-time analysis: sending a signal once a pattern is not recognized by the system
- Scrupulousness: end-to-end gap-free monitoring to identify smallest anomalies
- Accuracy: avoiding nuisance alerts and false positives/negatives triggered by static thresholds
- Self-learning: AI-driven algorithms learn from data patterns and deliver predictions or answers as required

The objective is not just to provide solution but to release the product (AI platform) which also shall lead to company organization transformation from outsource to product development model.

The target customers are companies from automotive testing industry.

There are different use cases for testing:

- Measurement device testing when Outliers between the tests (Time Series)
- Unit under the Test (UUT) testing with collection of measurement devices (Test Cell) when Outliers within the test (Batch and time series) and Outliers between the tests (Time Series)

The following challenges shall be addressed:

- Huge number of configuration parameters (input feature matrix)
- Validation of raw continuous data calculation

PART I TECHNICAL

CHAPTER 1 PREDICTIVE MAINTANANCE

1.1Background and significance of PdM

Predictive maintenance often refers to the use of data, analytics, and ML techniques to predict and prevent equipment failures or breakdowns before they occur. It involves continuously monitoring the condition and performance of equipment or systems and leveraging historical data to forecast potential failures or maintenance needs.

The background of predictive maintenance can be traced back to the evolution of maintenance strategies in industries. Traditionally, maintenance practices were primarily reactive or preventive. Reactive maintenance involved fixing equipment only after it failed, leading to unplanned downtime and costly repairs. Preventive maintenance, on the other hand, involved following a fixed schedule for maintenance tasks, such as replacing parts or conducting inspections, regardless of the actual condition of the equipment.

The limitations of reactive and preventive maintenance led to the development of predictive maintenance as a more proactive and data-driven approach. The increasing availability of sensors, IoT devices, and advancements in data analytics enabled organizations to collect vast amounts of data from equipment, systems, and processes. This data can include variables such as temperature, vibration, pressure, energy consumption, and more.

The significance of predictive maintenance lies in its potential to transform maintenance practices and deliver several key benefits:

- Cost Reduction: By identifying potential failures in advance, predictive maintenance helps organizations avoid costly unplanned downtime, emergency repairs, and unnecessary preventive maintenance activities. This can result in significant cost savings in terms of reduced equipment downtime, minimized maintenance labor and spare part costs, and optimized inventory management.
- Increased Equipment Reliability: Predictive maintenance enables organizations to detect and address issues early, preventing major breakdowns and extending the lifespan of equipment. By proactively addressing maintenance needs, organizations can improve the reliability and availability of critical assets, leading to improved operational efficiency and customer satisfaction.
- Improved Safety and Risk Management: Predictive maintenance helps identify potential safety hazards or malfunctions in equipment, allowing for timely corrective actions. It enhances worker safety by minimizing the likelihood of accidents caused by equipment failures. Additionally, by mitigating the risk of failures and breakdowns, organizations can reduce the potential environmental impact or regulatory non-compliance associated with equipment malfunctions.
- Enhanced Asset Performance: Through continuous monitoring and analysis of equipment data, predictive maintenance provides insights into equipment performance, degradation patterns, and usage patterns. This information can be leveraged to optimize maintenance schedules, improve asset utilization, and make data-driven decisions regarding repairs, replacements, or upgrades.

- Data-Driven Decision Making: Predictive maintenance relies on advanced analytics and machine learning algorithms to process and analyze large volumes of data. The insights generated from these analyses can support data-driven decision making in maintenance strategies, resource allocation, spare parts management, and overall asset management.

Overall, predictive maintenance has gained significant attention and adoption across various industries due to its potential to transform maintenance practices from reactive or preventive to proactive and data-driven. The ability to anticipate failures, reduce downtime, optimize maintenance activities, and enhance overall operational efficiency makes it a crucial aspect of modern asset management and operational excellence.

1.2Importance of anomaly detection in predictive maintenance

Anomaly detection plays a critical role in the field of predictive maintenance. Here are several reasons highlighting the importance of anomaly detection in predictive maintenance:

- Early Fault Detection: Anomaly detection techniques help identify deviations from normal behavior in equipment or system data. By detecting anomalies early, organizations can identify potential faults or malfunctions before they escalate into major failures. This early detection allows for proactive maintenance actions to be taken, minimizing downtime and reducing the risk of costly repairs or replacements.
- Improved Maintenance Planning: Anomaly detection provides insights into the condition and performance of assets. By monitoring data for anomalies, organizations can prioritize maintenance activities based on

the severity or criticality of detected anomalies. This enables more efficient maintenance planning and resource allocation, optimizing the use of maintenance personnel, spare parts, and other resources.

- Reduced Downtime and Cost: Anomalies often indicate equipment degradation, impending failures, or abnormal operating conditions. By detecting anomalies in real-time or near real-time, organizations can take immediate corrective actions to prevent unexpected downtime. This proactive approach reduces the impact of equipment failures on production or service delivery, leading to significant cost savings and improved operational efficiency.
- Enhanced Equipment Reliability: Anomaly detection helps identify patterns or trends that could indicate early signs of equipment degradation or performance decline. By continuously monitoring data for anomalies, organizations can detect and address these issues proactively. This results in improved equipment reliability, increased availability, and extended asset lifespan.
- Condition-Based Maintenance: Anomaly detection enables organizations to move from time-based or fixed-schedule maintenance to condition-based maintenance strategies. Instead of performing maintenance tasks based on predetermined schedules, maintenance actions are triggered based on the detection of anomalies or changes in asset behavior. This approach optimizes maintenance efforts, reduces unnecessary maintenance activities, and minimizes costs while ensuring equipment reliability.
- Data-Driven Decision Making: Anomaly detection leverages data analytics and machine learning algorithms to analyze large volumes of

equipment data. The insights generated from anomaly detection can support data-driven decision making in maintenance strategies, resource allocation, and asset management. By using data to guide decisions, organizations can optimize maintenance activities, improve operational efficiency, and reduce costs.

- Safety and Risk Mitigation: Anomalies in equipment data may indicate potential safety hazards or abnormal operating conditions. Detecting and addressing these anomalies promptly helps organizations ensure the safety of workers, mitigate risks associated with equipment failures, and comply with safety regulations. Anomaly detection contributes to maintaining a safe working environment and minimizing the potential for accidents or incidents caused by faulty equipment.

1.3 Anomaly detection of predictive maintenance in automotive testing

Anomaly detection is an essential component of predictive maintenance in automotive testing. It helps identify abnormal behavior or deviations from expected patterns during testing processes, enabling early detection of potential issues. Here are some key aspects of anomaly detection for predictive maintenance in automotive testing:

- Test Data Analysis: Anomaly detection techniques are applied to the data collected during various automotive testing processes, including durability testing, performance testing, reliability testing, and environmental testing. By analyzing test data, anomalies can be identified, indicating unexpected behavior, deviations from standard test results, or signs of potential failures.

- Failure Prediction: Anomaly detection is used to predict potential failures or malfunctions based on test data analysis. By monitoring and analyzing data from sensors, control systems, and performance metrics, anomalies can be detected, indicating deviations from expected behavior or signs of impending failures. Early detection of anomalies enables proactive measures to be taken, preventing unexpected failures during testing.
- Quality Control: Anomaly detection plays a crucial role in quality control during automotive testing. By monitoring test data for anomalies, it becomes possible to identify issues such as manufacturing defects, faulty components, or abnormal behavior that may impact the quality of the tested vehicles. Anomalies in the test data trigger alerts, enabling quick investigation and corrective actions to ensure the reliability and safety of the tested vehicles.
- Test Process Optimization: Anomaly detection helps optimize the automotive testing process by identifying opportunities for improvement. By analyzing test data, anomalies can be detected, indicating inefficiencies, deviations from standard procedures, or areas where the testing process can be enhanced. These insights enable process optimization, leading to increased efficiency, reduced testing time, and improved resource utilization.
- Test Equipment Monitoring: Anomaly detection techniques are utilized to monitor the health and performance of testing equipment, such as dynamometers, sensors, or data acquisition systems. By continuously monitoring data from these equipment, anomalies can be detected, indicating issues such as calibration errors, sensor drift, or malfunctions.

Early detection of anomalies enables timely maintenance or replacement, ensuring the accuracy and reliability of the testing equipment.

- Test Validation and Verification: Anomaly detection helps in the validation and verification of automotive testing processes. By comparing test results against expected benchmarks or reference data, anomalies can be detected, indicating discrepancies or unexpected variations. Anomalies in test results trigger investigations to ensure the accuracy and reliability of the testing process and to validate the conformity of the tested vehicles.
- Decision Support Systems: Anomaly detection supports decision-making in automotive testing by providing insights into potential issues or anomalies. By integrating anomaly detection algorithms into decision support systems, testing teams can make informed decisions regarding test results, equipment maintenance, process optimization, and quality control. These systems enable efficient resource allocation, timely interventions, and proactive measures for ensuring the accuracy and effectiveness of the automotive testing process.

CHAPTER 2 ANOMALY TYPES AND ALGORITHMS

Anomaly detection involves identifying the differences, deviations, and exceptions from the norm in a dataset. It’s sometimes referred to as outlier detection (i.e., looking at a dataset to identify any outlying or unusual datapoints, data groups, or activity).

2.1Different Types of Anomalies

Not all anomalies are equal. In fact, they can be split into three broad categories:

- Point anomalies. A point anomaly is where a single datapoint stands out from the expected pattern, range, or norm. In other words, the datapoint is unexpected. See Figure 2 Examples of point anomalies
- Collective anomalies. A collective anomaly occurs where single datapoints looked at in isolation appear normal. When you look at a group of these datapoints, however, unexpected patterns, behaviors, or results become clear. See Figure 3 An irregular heart beat is an example of a collective anomaly.
- Contextual anomalies. Instead of looking at specific datapoints or groups of data, an algorithm looking for contextual anomalies will be interested in unexpected results that come from what appears to be normal activity. The crucial element here is context: Are the results out of context? See Figure 4 Example of contextual anomaly detection using the Twitter AnomalyDetection package in R

2.2Univariate/Multivariate

Univariate and multivariate anomalies are two different types of anomalies based on the dimensionality of the data being analyzed. Here's an explanation of each:

- Univariate Anomalies: Univariate anomalies focus on the analysis of a single variable or feature at a time. In this case, anomalies are identified by comparing the values of a single variable against a defined threshold or statistical measure. The analysis considers only one aspect of the data and looks for deviations from the expected behavior for that particular variable. For example, if you are examining the temperature data for a city, univariate anomalies would involve identifying temperature readings that are significantly higher or lower than the average or fall outside a predefined range.

- Multivariate Anomalies: Multivariate anomalies, on the other hand, consider the relationships and interactions between multiple variables simultaneously. Instead of examining individual variables in isolation, these anomalies are detected by analyzing the joint behavior of multiple variables together. The anomalies are identified based on deviations from the expected behavior of the entire multivariate dataset, considering the interactions between variables. For instance, if you are analyzing sales data, multivariate anomalies could involve identifying combinations of variables, such as high sales volume, low price, and unusual customer behavior occurring together, which are not typically observed.

2.3Learning type

Supervised learning is the scenario in which the model is trained on the labeled data, and trained model will predict the unseen data. Whereas in unsupervised learning, no labels are presented for data to train upon. Each of the methodologies has its advantages, and disadvantages like supervised learning models do produce highly accurate results, whereas unsupervised learning models do not improve performance over the period. One thumb rule could be, whenever any point is falling beyond 99th percentile value, that data point can be classified as an anomaly using unsupervised learning, but this method is too trivial.

There is a third methodology called semi-supervised learning, which is the combination of both supervised and unsupervised learning, here we would be explaining with the application using time series data in detail. See Figure 5

Example of semi-supervised learning on timeseries data for Anomaly detection, where Regression model (like Random forest regressor or XGBoost regressor) can be used to train on the past 10 days to predict next 2 days based on various explanatory variables like hour flag, which day of the week flag, some lag variables, etc. Once the actual values are predicted from the model, upper and lower bounds are calculated based on the standard deviation observed from the prediction of the model for 2 days. Quantity to be added in either direction to the actual values is computed by 1.96 * standard deviation etc. The upper limit is the predicted central value + 1.96 * standard deviation, and the Lower limit is the predicted central value – 1.96 * standard deviation.

Once the upper and lower bounds are calculated, it will be overlaid with actual values and will highlight the anomaly whenever the real value do breach either upper or lower limits. One can say why this is a semi-supervised learning model is

- A model trained on actual data is like a supervised regression model, which is only the first part of the whole process.
- In the second stage, Anomalies predicted (1 represents an anomalous data point, and 0 is non-anomalous data point) whenever actual value breaches either upper or lower bounds is a representation of unsupervised learning model, where the multiplier values are fixed (like 1.96,2.56, etc.). Only variable in this phase is standard deviation which is not in control for model

In this way, semi-supervised learning is utilized for prediction of the anomalous data points on time-series data. Even though the second phase of the model is unsupervised, where bounds do not learn with the data. However, the user can manually reset the values to adjust to the signal accordingly. Let us say if the limits are too tight, and through too many anomalies, the user needs to increase the bands like changing the constant from 1.96 to 2.56. Similarly, if the bounds are too broad and every value is getting camouflaged, then the user needs to tighter the limits from 1.96 to 1.63, etc. to make bounds adjust to the signal type.

2.3.1 Supervised algorithms

Supervised algorithms can also be employed for anomaly detection. While unsupervised learning is commonly used for anomaly detection, supervised algorithms can offer advantages in certain scenarios.

One approach is to treat anomaly detection as a binary classification problem, where labeled data points are available, including both normal and anomalous instances. The supervised algorithm can then be trained using these labeled examples to learn the patterns and characteristics of normal data. Once trained, the algorithm can classify new instances as either normal or anomalous based on the learned patterns.

Some popular supervised algorithms for anomaly detection include:

- Support Vector Machines (SVM): SVMs can be used for binary classification, separating normal instances from anomalies based on a defined decision boundary. Anomalies that fall on the wrong side of the decision boundary are classified as outliers.
- Random Forests: Random Forests are ensemble learning algorithms that can be utilized for anomaly detection. By training a forest of decision trees on labeled data, anomalies can be identified based on the disagreement among individual trees in the forest.

- Neural Networks: Neural networks, particularly deep learning models, can be used for supervised anomaly detection. By training a neural network on labeled data, the network learns to recognize normal patterns and can flag instances that deviate significantly from those patterns as anomalies.

It's important to note that supervised anomaly detection methods require labeled data, which can be a limitation as obtaining labeled anomaly data may be challenging or costly. Additionally, these methods may struggle with detecting novel or previously unseen anomalies since they rely on patterns learned from labeled training data.

2.3.2 Unsupervised algorithms

Unsupervised algorithms for anomaly detection are machine learning techniques that aim to identify anomalies or outliers in a dataset without the need for labeled training data. These algorithms learn the normal patterns or behavior of the data and flag instances that deviate significantly from those patterns as anomalies. Here are a few commonly used unsupervised algorithms for anomaly detection:

- Density-Based Methods: Density-based algorithms, such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise), identify anomalies as data points that fall in sparser regions of the data distribution. They classify points in dense areas as normal and points in sparse areas as anomalies.
- Clustering-Based Methods: Clustering algorithms, such as k-means or Gaussian Mixture Models (GMM), group similar data points together based on certain similarity measures.

- Distance-Based Methods: Distance-based algorithms, such as k-nearest neighbors (k-NN) or Local Outlier Factor (LOF), detect anomalies by measuring the distance between a data point and its neighboring points. Anomalies are identified as data points that have significantly larger distances to their neighbors compared to the majority of the data.
- Statistical Methods: Statistical algorithms, such as Gaussian distribution modeling or the Z-score method, assume that the normal data follows a specific statistical distribution. Anomalies are identified based on statistical measures, such as deviations from the mean or exceeding a certain threshold.
- Autoencoders: Autoencoders are neural network-based models that aim to reconstruct the input data. Anomalies are identified as instances that have higher reconstruction errors compared to the majority of the data. Autoencoders learn the representation of normal patterns and struggle to accurately reconstruct anomalous instances.

2.4Decision type

AI systems and ML-based solutions offer the advantage of continuous learning and improvement, resulting in better and more precise results over time. The learning process typically involves several stages, combining automated processes with human assistance. Here is a breakdown of the stages:

- Datasets are fed to an AI system: The AI system is provided with relevant datasets containing examples of normal behavior and known anomalies. These datasets serve as training data for the system to learn and establish patterns.

- Data models are developed based on the datasets: Using the provided datasets, the AI system develops data models or algorithms that capture the patterns and characteristics of normal behavior. These models form the basis for identifying anomalies in future data.
- A potential anomaly is raised: As the AI system encounters new data or transactions, it compares them against the established data models. If a transaction deviates significantly from the learned patterns, it is flagged as a potential anomaly.
- A domain expert approves the deviation: To ensure accuracy, a domain expert or human reviewer examines the flagged deviation and determines if it indeed represents an anomaly. This step adds a layer of human expertise and judgment to the process.
- The system learns from the action and builds upon the data model: Once the deviation is confirmed as an anomaly, the AI system incorporates this feedback into its learning process. It updates the data model, incorporating the newly identified anomaly to improve future predictions.
- The system continues to accumulate patterns: Over time, the AI system continues to accumulate patterns based on the preset conditions and the feedback received from domain experts. As more anomalies are identified and confirmed, the system refines its understanding of normal behavior and becomes more accurate in detecting anomalies.

By going through these stages iteratively, the AI system becomes increasingly proficient in identifying anomalies and generating precise results. The combination of automated learning from data and human expertise ensures the continuous improvement and effectiveness of the AI system in anomaly detection.

CHAPTER 3 PROPOSED MODEL/METHODS

3.1Multivariate, semi-supervised, AutoEncoder

he autoencoder method for anomaly detection is an approach that utilizes autoencoders, which are neural network architectures, to identify anomalies or abnormalities in a dataset. Autoencoders are unsupervised learning models that aim to reconstruct the input data in the output layer, using a compressed representation in an intermediate layer called the latent space.

Here's how the autoencoder method for anomaly detection works:

- Training Phase: In the training phase, the autoencoder is trained on a dataset consisting of normal or non-anomalous instances. The input data is passed through the encoder part of the network, which maps the data to the latent space, reducing its dimensionality. The decoder part of the network then attempts to reconstruct input data from the latent space representation. During training, the autoencoder learns to capture the underlying patterns and regularities in the normal data.
- Reconstruction Error: The reconstruction error, which represents the discrepancy between the input data and its reconstructed output, is computed during the training phase. The autoencoder aims to minimize this error, learning to generate accurate reconstructions of the normal instances. The reconstruction error serves as a measure of how well the autoencoder can reproduce the normal patterns present in the training data.
- Anomaly Detection: Once the autoencoder is trained on normal data, it can be used for anomaly detection. During the testing phase, input data

instances are fed into the trained autoencoder. If an input instance deviates significantly from the normal patterns learned by the autoencoder, it will result in a higher reconstruction error. Instances with high reconstruction errors are considered potential anomalies.

- Thresholding: To determine whether an instance is anomalous or not, a threshold is set on the reconstruction error. Instances with reconstruction errors exceeding the threshold are classified as anomalies, while those below the threshold are considered normal.

Please see Figure 6 Autoencoder network

The Multivariate, semi-supervised AutoEncoder method for anomaly detection offers several benefits that make it a valuable approach in various applications. Here are some advantages of this method:

- Comprehensive Anomaly Detection: By considering multiple variables or features simultaneously, the Multivariate AutoEncoder method can detect complex anomalies that involve interactions and dependencies among different variables. This allows for a more comprehensive analysis of the data and a better understanding of abnormal patterns.
- Unsupervised Learning: The AutoEncoder operates in an unsupervised learning manner, meaning it doesn't require labeled anomaly data during the training phase. This is advantageous because labeled anomaly data can be scarce or expensive to obtain. The model learns from normal data, enabling it to detect anomalies in unseen data during the testing phase without relying on pre-labeled anomalies.
- Dimensionality Reduction: AutoEncoders are capable of reducing the dimensionality of the data, which is beneficial in anomaly detection. By

compressing the data into a lower-dimensional latent space, the method can capture essential information and discard noise or irrelevant features. This aids in identifying relevant patterns and anomalies in a more efficient manner.

- Nonlinear Relationships: Unlike linear techniques such as PCA (Principal Component Analysis), AutoEncoders can capture nonlinear relationships between variables. This flexibility allows them to represent complex data distributions more accurately, enabling the detection of anomalies that involve nonlinear dependencies among features.
- Reconstruction Error as Anomaly Indicator: The reconstruction error, which measures the dissimilarity between the input data and the reconstructed output, serves as an anomaly indicator. Higher reconstruction errors suggest a higher likelihood of anomalies. This simplicity makes it easy to set a threshold for anomaly detection and interpret the results.
- Semi-Supervised Learning: Incorporating a small amount of labeled anomaly data in the training process makes the method more adaptable to specific anomaly patterns or characteristics. It allows the AutoEncoder to learn from both normal and labeled anomaly data, improving its ability to generalize and detect anomalies accurately.
- Real-Time Anomaly Detection: The Multivariate, semi-supervised AutoEncoder method can be deployed in real-time anomaly detection systems. It can process data in near real-time, making it suitable for applications where timely detection and response to anomalies are critical, such as industrial systems, cybersecurity, and financial fraud detection.

- Adaptability to Various Domains: The method is versatile and can be applied to different domains and types of data, including time series data, sensor data, images, text, and more. This adaptability makes it applicable across various industries, such as manufacturing, finance, healthcare, and cybersecurity.

CHAPTER 4 IMPLEMENTATION/EXPERIMENTAL CHAPTER

Project code is attached to the document. Please refer to Appendix Project Code

4.1Data collection

Data collection and preprocessing are crucial steps in anomaly detection to ensure the quality and reliability of the data used for analysis. Here are the key considerations for data collection and preprocessing in anomaly detection:

- Identify Relevant Data Sources: Determine the data sources that are relevant to the anomaly detection task. This could include sensor data, log files, transaction records, network traffic data, or any other data that is indicative of normal or abnormal behavior in the system.
- Data Collection: Collect the data from the identified sources. Depending on the nature of the data, this could involve extracting data from databases, accessing APIs, setting up data pipelines, or employing data acquisition systems. Ensure that the collected data represents a wide range of normal and abnormal scenarios to build a robust anomaly detection model.
- Data Cleaning: Clean the collected data to remove any inconsistencies, errors, or missing values that could adversely impact the accuracy of the anomaly detection model. Handle missing data through imputation techniques or by removing incomplete instances, depending on the extent of missingness and the specific requirements of the analysis.
- Feature Selection and Engineering: Analyze the collected data and select relevant features or variables that are likely to contain meaningful information for detecting anomalies. Feature engineering techniques may

be employed to transform or create new features that better capture the underlying patterns or relationships in the data. This could involve statistical transformations, time-based aggregations, or domain-specific knowledge.

- Data Scaling and Normalization: Scale and normalize the features to ensure they have similar ranges or distributions. This step is particularly important when using distance-based anomaly detection algorithms that rely on the similarity between instances. Common techniques include z-score normalization, min-max scaling, or robust scaling.
- Handling Imbalanced Data: Anomaly detection datasets often have a class imbalance, where the number of normal instances far exceeds the number of anomalies. Consider techniques such as oversampling, under sampling, or synthetic data generation to balance the dataset, ensuring that anomalies are adequately represented during training.
- Outlier Removal: Identify and handle any outliers that are not genuine anomalies but rather data artifacts or measurement errors. Outliers can skew the distribution and impact the performance of anomaly detection models. Techniques like the interquartile range (IQR), z-score, or domain knowledge can be used to identify and handle outliers appropriately.
- Train-Test Split: Split the preprocessed data into training and testing datasets. The training dataset is used to build the anomaly detection model, while the testing dataset is used to evaluate the performance of the model on unseen data. Consider the temporal aspect of the data when splitting to ensure that the testing data represents future time periods or unseen instances.

- Additional Preprocessing Techniques: Depending on the specific characteristics of the data and the chosen anomaly detection algorithm, additional preprocessing steps may be required. These could include dimensionality reduction techniques (e.g., PCA) to reduce computational complexity, data discretization, or time series preprocessing techniques (e.g., smoothing, differencing).

The following specific actions were done:

- Simulated but not real dataset for one test for specific configuration was exported from database to Excel.
- Dataset has been anonymized
- Data was transformed and read to Pandas dataframe in Jupyter notebook
- Whole NA features were dropped; some NA values were replaced by previously valid ones.
- Final Data frame includes 71 numeric continuous features with no outliers which can be treated as time series. See Figure 7 numeric continuous features of the dataset

4.2Data processing and analysis

Original dataset was perturbed with 3.6% outliers from total number. See Figure 8 Dataset was perturbed with 3.6% outliers from total number

Exploratory data analysis shows labeled outliers. See Figure 9 Exploratory data analysis shows labeled outliers in perturbed dataset

We use a 3 layer neural network: First layer has 10 nodes, middle layer has 2 nodes, and third layer has 10 nodes. We use the mean square error as loss function, and train the model using the "Adam" optimizer. See Figure 10 AutoEncoder neural network model

Train model for 100 epochs, batch size of 10.

4.3Verification of the suitability of findings

Fitting the model: To keep track of the accuracy during training, we use 5% of the training data for validation after each epoch (validation_split = 0.05)

By graphing the distribution of the calculated loss within the training set, one can use this to identify a suitable threshold value for identifying an anomaly. In doing this, one can make sure that this threshold is set above the "noise level", and that any flagged anomalies should be statistically significant above the noise background. See Figure 11 Threshold selection

From the above loss distribution, let us try a threshold of 0.3 for flagging an anomaly. We can then calculate the loss in the test set, to check when the output crosses the anomaly threshold.

We then calculate the same metrics also for the training set, and merge all data in a single dataframe

4.4Analysis, interpretation and summary of the findings

Outliers were predicted then (see Figure 12 Outlier prediction) with accuracy F1 score: 0.5 -- Accuracy: 0.97 -- Recall: 0.35. See Figure 13 AutoEncoder accuracy

Accuracy can be improved either by fine tuning AutoEncoder model or by applying another AI algorithm.

CHAPTER 5 TECHNOLOGY CONCLUSION/RECOMMENDATIONS

AI algorithms like AutoEncoder allows to find out outliers what is required for predictive maintenance use cases.

However, its necessary to invest in every of the following areas to get accurate and reliable results:

- Having the right data available (Relevant, sufficient, Quality)
- Framing the problem appropriately (Feature engineering)
- Evaluating the predictions properly (Model evaluation)

KSE | Business Education

PART II BUSINESS

CHAPTER 6 BUSINESS GOALS AND DIAGNOSTICS

6.1Company strategy

Company strategical goal is to switch from outsourcing to product business model for B2B segment to increase company business marginality and to gain intangible assets for new opportunities.

The following scenario are considered:

- Scenario 1 (Base): New product sales via partner company who has own developed sales channels
- Scenario 2: To build up own marketing and sales process for new product sales
- Scenario 3: New product sale to an investor
- Scenario 4 (Plan B): To sell new product as outsourcing hours to partner but with higher rate

6.2Internal SWOT analysis

Our company Strength, Weaknesses can be seen in Figure 14 Company internal SWOT analysis

6.3Organizational behavior transformation

The following Intangible assets shall be acquired and human capital transformations shall occur for changing from outsourcing to product development model:

- Institualization of innovation and creativity culture

- Getting product development skills and mindset
- Acquiring new hard skills like data science, marketing, sales, finance etc
- The following tools can help to achieve that:
- Management support, communication and motivation
- Company process transformation (Introduction Design Thinking etc.)
- Alignment with HR hiring, retention and reword process with focus on soft skills
- Getting new talents with required skills and both external and internal trainings, conferences, meetups etc.

6.4Adaptability to the realities of a changing

The following measures were done or planned to adopt to the influence of the war:

- Start of investment is delayed to 2024: homework shall be done before.
- Esurance of the safety of employees: keep the key Team.
- Maintain a stable infrastructure: electricity, Internet, data migration to cloud.
- Build strong partnerships: to work more closely with key partner and to try to find new ones.
- Emphasize customer support.
- Diversify revenue streams: company representative in Germany to have direct sales in EU.
- Seek government assistance: booking of key employees, business trips allowance.

CHAPTER 7 MARKETING STRATEGY

7.1Market Micro-environmental audit

Artificial intelligence (AI) is rapidly transforming various industries, and the automotive industry is no exception. One of the areas where AI is making a significant impact in the automotive industry is predictive maintenance. Predictive maintenance is the use of AI-powered tools and algorithms to predict when a vehicle is likely to experience a failure or malfunction and to proactively address the issue.

First, let's look at the benefits of using AI for predictive maintenance in the automotive industry. One of the most significant benefits is that it can reduce maintenance costs by up to 30% by identifying and addressing issues before they become major problems. This is because AI-powered tools can detect potential issues with greater accuracy than human inspection.

Additionally, predictive maintenance can minimize vehicle downtime and increase safety for drivers and passengers by addressing issues proactively.

ROI: The return on investment (ROI) for predictive maintenance can be as high as 10:1, meaning that for every dollar invested in predictive maintenance, companies can expect to see $10 in return. (Source: Deloitte)

7.1.1 Market size

The automotive PdM market is expected to achieve a CAGR of ~28% from 2019 and 2027.

See Figure 15 Market Micro-environmental audit for automotive predictive maintenance. The market is projected to be driven by the increasing demand for automotive vehicles all over the world.

The global automotive PdM market is expected to reach about US$ 2.7 Bn by 2027, from ~US$ 390 Mn in 2019. Increased focus on reducing unplanned maintenance downtime for vehicles is boosting the automotive predictive maintenance market.

In conclusion, AI is transforming the industry in the area of predictive maintenance. By using AI-powered tools and algorithms, automotive companies can reduce maintenance costs, minimize vehicle downtime, and increase safety for drivers and passengers. While there are challenges in implementing these tools, the benefits are significant, and the automotive industry is rapidly adopting these technologies. As AI continues to evolve, so we can expect even more innovative solutions in the area of predictive maintenance in the automotive industry.

7.1.2 Market Drivers

The following factors are driving the growth of the predictive maintenance market in the automotive industry:

- Increasing demand for vehicle safety and reliability
- Growing need for reducing maintenance costs and downtime
- Rising adoption of Industry 4.0 and IoT technologies
- Increasing focus on reducing carbon emissions and overall sustainability
- Growing need for optimized vehicle performance and improved fuel efficiency

7.1.3 Market Segmentation

The predictive maintenance market can be segmented based on component, deployment mode, application, and region.

- Component: into services and solutions. Solutions segment is expected to hold a larger share due to the increasing adoption of software tools for predictive maintenance.
- Deployment Mode: into on-premises and cloud-based. The cloud-based segment is expected to grow at a faster rate due to its flexibility and scalability.
- Application: Manufacturing: This application segment includes predictive maintenance solutions that are used in the production process of vehicles, such as assembly line equipment and robotic systems. Predictive maintenance solutions can help prevent equipment downtime and reduce maintenance costs. Fleet management: This application segment includes predictive maintenance solutions that are used to monitor and maintain large fleets of vehicles, such as those used in logistics and transportation companies. Predictive maintenance solutions can help optimize vehicle uptime, reduce maintenance costs, and improve safety.Aftermarket services: This application segment includes predictive maintenance solutions that are used by automotive service providers to offer preventive maintenance services to their customers. Predictive maintenance solutions can help service providers offer more effective and efficient services to their customers, which can improve customer satisfaction and retention.Warranty and insurance: This application segment includes predictive maintenance solutions that are used by automotive manufacturers and insurance companies to offer warranty and insurance

services to their customers. Predictive maintenance solutions can help manufacturers and insurers reduce costs and risks associated with warranty and insurance claims, while also improving customer satisfaction..

- Region: into NA, Europe, Asia Pacific, and Rest of the World. Asia expected to grow at the highest CAGR because of increasing adoption of Industry 4.0 and IoT technologies in the region.

Our target customers are Automotive OEMs in Europe which needs on-premise solutions for manufacturing application.

7.2Competition Micro-environmental audit

Main competitors are big companies who have a lot of resources on one hand. But on another hand their products are general purpose and don't provide specific customizable solutions.

The following are some of the major players operating in the predictive maintenance market in the automotive industry:

- IBM Corporation
- Microsoft Corporation
- SAS Institute Inc.
- PTC Inc.
- Software AG
- SAP SE
- General Electric
- Rockwell Automation, Inc.
- Schneider Electric SE
- Hitachi Ltd.

See Figure 16 Predictive maintenance company ranking

7.3Challenges for successful adoption of prediction maintenance

There are several technical and business challenges that need to be addressed when proposing a new product for predictive maintenance in the automotive industry. Some of these challenges include:

- Data quality and availability: One of the biggest challenges in predictive maintenance is ensuring that the data used to train and validate models is accurate and representative of real-world conditions. Data collection and quality assurance can be time-consuming and expensive, particularly if multiple data sources are involved.
- Integration with existing systems: Most automotive OEMs have multiple legacy systems and platforms that need to be integrated with any new predictive maintenance solution. This can be a complex and time-consuming process, requiring significant investment in IT resources.
- Scalability and reliability: Predictive maintenance solutions need to be scalable to handle large volumes of data and able to operate reliably in harsh environments. This can be particularly challenging in the automotive industry, where equipment is exposed to extreme temperatures, humidity, and vibration.
- Cost-effectiveness: Predictive maintenance solutions need to be cost-effective, providing a clear return on investment for the OEM. This can be challenging if the solution requires significant investment in hardware, software, or personnel.

- Competitive landscape: The predictive maintenance market is becoming increasingly crowded, with many established players already offering solutions to the automotive industry. Differentiating a new product from the competition and demonstrating its value proposition can be a significant challenge.

The successful implementation of a PdM program in the automotive industry requires a significant cultural change and human capital transformation. Here are some key points to consider:

- Training and upskilling: Predictive maintenance programs require skilled personnel who can operate and maintain the systems, analyze data, and make data driven decisions. Companies need to invest in training and upskilling their workforce.
- Collaboration: Successful predictive maintenance programs require collaboration between various departments within the company, including IT, engineering, and maintenance.
- Data-driven decision-making: PdM programs rely heavily on data to inform decision-making. Companies need to instill a data-driven culture, where decisions are made based on insights from data rather than intuition or experience.
- Performance metrics: Establishing clear performance metrics is essential for tracking progress and ensuring accountability. Companies need to develop KPIs that are aligned with business objectives and measure the effectiveness of the predictive maintenance program.
- Leadership buy-in: A successful predictive maintenance program requires strong leadership buy-in from top executives. Leaders need to

communicate the importance of the program, set clear goals and objectives, and provide the necessary resources and support to ensure its success.

- Change management: Implementing a predictive maintenance program involves significant change management. Companies need to engage with employees and stakeholders throughout the process, communicate the benefits of the program, and address any concerns or resistance to change.
- Continuous improvement: Finally, companies need to embrace a culture of continuous improvement. Predictive maintenance programs are not static, and companies need to be open to evolving their systems and processes as new technologies and best practices emerge.

In summary, cultural change and human capital transformation are critical for the successful implementation of a predictive maintenance program in the automotive industry. Companies need to invest in training and upskilling their workforce, foster collaboration and knowledge-sharing, establish clear performance metrics, secure leadership buy-in, manage change effectively, and embrace a culture of continuous improvement.

7.4Alternative maintenance options

Here is a list of alternative maintenance options for predictive maintenance, which can be used by our customers:

- Preventive Maintenance: Scheduled maintenance activities performed at predetermined intervals or usage thresholds to prevent failures.
- Reactive Maintenance: Repairing or replacing equipment components only after a failure occurs, also known as "breakdown" or "run-to-failure" maintenance.

- Condition-Based Maintenance (CBM): Monitoring the actual condition of equipment using sensors and data collection to detect anomalies and trigger maintenance actions.
- Reliability-Centered Maintenance (RCM): Systematically analyzing failure modes, determining maintenance strategies, and optimizing maintenance efforts based on equipment reliability, operational impact, safety risks, and costs.
- Run-to-Completion: Running equipment until it fails completely before performing any maintenance or replacement, typically used for non-critical assets.
- Predictive Analytics: Using advanced analytics and machine learning techniques to analyze data and predict future equipment failures or maintenance needs.
- Proactive Maintenance: Performing maintenance tasks based on proactive analysis and prediction of equipment behavior, combining elements of preventive and predictive maintenance.
- Prescriptive Maintenance: Utilizing analytics and algorithms to recommend specific maintenance actions based on real-time data and equipment conditions.
- Advanced Troubleshooting: Applying specialized techniques and expertise to diagnose and resolve complex equipment issues when they occur.
- Risk-Based Maintenance: Prioritizing maintenance activities based on risk assessment and focusing efforts on critical assets or failure modes with the highest impact.

- Reliability Improvement Programs: Implementing continuous improvement initiatives to enhance equipment reliability, reduce failure rates, and optimize maintenance practices.
- Asset Health Management: Monitoring and managing the overall health and performance of assets through integrated data collection, analysis, and maintenance strategies.

It's important to note that the selection of maintenance options depends on factors such as equipment criticality, reliability, cost implications, available resources, and the organization's specific needs and goals. Implementing a combination of these alternative maintenance options can help organizations optimize their maintenance practices and improve overall equipment performance and reliability.

Possible criteria for maintenance option selection are presented in Figure 17 Alternative maintenance options

7.5Value proposition

There are several motivations for big automotive OEMs to buy predictive maintenance solutions from small vendors like we, including:

- Innovation: to have more flexibility and agility to innovate and develop cutting-edge technology solutions. This can be attractive to big automotive OEMs who are looking to stay ahead of the curve and differentiate themselves in the market.
- Customization: to offer more customization options and tailor solutions for each individual customer. This can be appealing to big automotive OEMs who have unique requirements that may not be met by off-the-shelf solutions.

- Cost: to offer more affordable solutions than larger, more established vendors. This can be attractive to big automotive OEMs who are looking to reduce costs and improve their bottom line.
- Expertise: to have a deep understanding of particular niche or specialty. This can be attractive to big automotive OEMs who are looking for specific expertise in a particular area.
- Partnership Opportunities: Big automotive OEMs may see small vendors as potential partners in the development of new products and services. By working closely with small vendors, big automotive OEMs can leverage their expertise and technology to create new solutions that meet the evolving needs of their customers.
- Consulting and Training: To provide services for training of OEMs personal and helping them with organization transformation for successful adoption of predictive maintenance.

In summary, big automotive OEMs may be motivated to buy predictive maintenance solutions from small vendors due to their innovation, customization options, affordability, expertise, speed to market, and potential for partnership opportunities.

7.6Marketing Mix

It is suggested to lunch new B2B IT product “Testing Intelligence (TI)” which is AI platform for predictive maintenance in measurement systems for automotive testing.

The target audience (macro segmentation) for the product are automotive companies who are suppliers of testing solutions for car manufacturers (OEMs). Micro segmentation target audience are such decision makers like head of developments, product managers, sales managers of corresponding companies who could empower own products and solutions with TI.

The TI product will have different subscription models depending on functional and service options (differentiated marketing).

The positioning statement of the product is "For (whom) automotive professionals (statement of need or opportunity) looking to reduce testing costs and time, (product) TI is the (category) AI platform for predictive maintenance that (point of difference) delivers state-of-the-art AI technology with proven and reliable results because (reason to believe) we are the experts in the both automotive testing and AI areas".

In fact, automotive industry is quite conservative regarding innovation implementation in general and particular for AI practical usage, that is why the key message is to build trust that our new product can already provide "state-of-the-art AI technology with proven and reliable results" as its tailored for specific automotive testing segment needs in comparison to AI solutions from other competitors who mainly provide either generic solutions or highly customized and therefore expensive ones. Also we as product supplier have necessary expertise and experience.

The marketing mix will include the following aspects taking into account that IT product is not tangible and it is provided for B2B segment.

Product:

- Superiority quality for performance and reliability but don't promise what you cannot deliver.
- Give consumers more information about the product, making the choice of the product better for them;

- Customer support service as a subscription option.
- Limited maintenance warranty

Price:

- Inelastic demand
- Going rate pricing strategy

Place:

- Direct distribution via virtual channel (internet)

Physical evidence:

- User Experience design of product GUI components and official website.

People:

- Key elements, requires professional recruitment, investing in staff & training, empowerment of staff, internal marketing

Process:

- It is about developing ways of delivering the product and related service that will add value to the customer experience
- Support service and Hotline

Promotion (communication mix):

- Advertising – Digital media: Context media SEO and Social media (LinkedIn, Xing) targeted ads
- Sales Promotion - Free trial period
- Personal Selling - Personal presentation and demonstration of products to end customer in thematic automotive testing/IT conferences, webinars, roadshows etc.

- Public Relations – articles and blog posts in own website, social media and high rating automotive and IT magazines (influencers) etc. to show expertise and benefits, communicating list of customers and related testimonials.
- Direct Marketing – Email marketing via LinkedIn contacts, Relationship Marketing for cross-selling and up-selling.

To ensure right investment allocation, promotional mix needs to be explained with Communications Strategy as the next step.

All components of the marketing mix have interdependencies and mutually influence each other. Together, they form an integral part of a company's business plan and, if effectively managed, can contribute to its tremendous success. The marketing mix necessitates a deep understanding of the market, extensive market research, and collaboration with various stakeholders, including customers, trade partners, manufacturing experts, and others.

CHAPTER 8 FINANCIAL MODELING

Financial modeling is presented for Scenario 1 (Base): New product sales via partner company who has own developed sales channels

8.1Business Model Canvas

Key Partners:

Plan is to involve our current customer as partner who has brand awareness for niche market and good relationships and available sales channels for target customers.Partner will perform marketing and sales activities.

Motivation for partnership: acquisition of particular resources and activities.

Key Activities:

Our Value Propositions require data science R&D (AI part) and software development (Platform part) as main activities

CATEGORIES: AI Platform

Key Resources:

Value Propositions require mainly human and financial resources for above activities.

TYPES OF RESOURCES: Human, Financial, Intellectual (data, software licenses), Physical (hardware)

Value Propositions:

We deliver reliable automatic prediction of maintenance issues before real failure occurs.

Superiority quality for performance and reliability but don’t promise what you cannot deliver.

Limited maintenance warranty

Give consumers more information about the product, making the choice of the product better for them.

We are helping customer to reduce equipment downtown time and costs for manual search and analyze of the issues.

We could bundle our value proposition with customization (e.g. retraining AI model for customer specific data) and providing related services (Support service and Hotline).

We are satisfying customer needs of cost reduction and increase of equipment usage efficiency.

CHARACTERISTICS::Cost Reduction, Performance, Customization, Convenience/Usability

Customer Relationships:

Our main target customer segments expect personal assistance and ready to pay for that (CARE contract).

Also sometimes they want to have tailored solution for their specific needs so that ready to consider customization and co-creation.

Channels:

Promotion (communication mix) will include:

- Advertising – Digital media: Context media SEO and Social media (LinkedIn, Xing) targeted ads
- Sales Promotion - Free trial period
- Personal Selling - Personal presentation and demonstration of products to end customer in thematic automotive testing/IT conferences, webinars, roadshows etc.
- Public Relations – articles and blog posts in own website, social media and high rating automotive and IT magazines (influencers) etc. to show expertise and benefits, communicating list of customers and related testimonials.
- Direct Marketing – Email marketing via LinkedIn contacts, Relationship Marketing for cross-selling and up-selling.
- Personal Selling work the best for OEMs

Customer Segments:

B2B

The target audience (macro segmentation) for the product are automotive companies who are suppliers of testing solutions for car manufacturers (OEMs)

Micro segmentation target audience are such decision makers like head of developments, product managers, sales managers of corresponding companies who could empower own products and solutions with TI.

Niche Market

Cost Structure:

The most important costs are salaries of high qualified darta scientits and software engineers.

Human resources are most expensive. Hardware for development and testing is on the second place.

R&D is most expensive activity.

Value Driven (focused on value creation, premium value proposition)

Fixed Costs (salaries, rents, utilities)

Revenue Streams:

Customers are willing to pay for reliable automated prediction of required maintenance.

Currently they are paying either for manual search and analyze of the issue or have to deal with occurred failure when they lose time=money.

The TI product will have different subscription models depending on functional and service options (differentiated marketing).

TYPES: Subscription Fees

FIXED PRICING: Product feature dependent, Customer segment dependent

Inelastic demand

8.2 Investment highlights

Investment highlights can be seen in Figure 18 Scenario 1 Investment highlights

8.3 Profit and Loss statement, Income distribution statement

Profit and Loss, Income distribution statements can be seen on Figure 19 Scenario 1 Profit and Loss statement, Income Distribution

Earnings distribution is presented on Figure 20 Scenario 1 Earnings graph

8.4 Balance Sheet

Balance Sheet statement can be seen on Figure 21 Scenario 1 Balance Sheet

8.5 Cashflow Statement

Cashflow statement can be seen on Figure 22 Scenario 1 Cashflow statement

Cashflow distribution is presented in Figure 23 Scenario 1 Cashflow graph

CHAPTER 9 TOTAL COST OF OWNERSHIP FOR THE CUSTOMER

The total cost of ownership (TCO) for an AI product in the area of predictive maintenance for automotive testing can vary significantly based on factors such as the specific requirements of the OEM (Original Equipment Manufacturer), the scope of the AI solution, and the complexity of the implementation. While it's challenging to provide an exact TCO without detailed information, here are some potential cost components to consider in addition to the product license price:

- Data Acquisition and Preparation: The cost of acquiring and preparing the necessary data for training and testing the predictive maintenance models specific to customers vehicles. This may involve data collection from various sensors, databases, and systems within customers infrastructure. Data cleaning, preprocessing, and data integration efforts are also part of this cost component.
- Infrastructure and Deployment Costs: This encompasses the hardware and software infrastructure required to deploy and run the AI product. It includes the cost of servers, storage, networking, and cloud computing resources necessary for hosting and processing the AI models. Additionally, deployment costs may include custom integration with customers' existing systems and infrastructure.
- Integration and Compatibility: The cost of integrating the AI product into customers existing automotive testing infrastructure, including compatibility testing, system integration efforts, and potentially adapting the AI solution to work seamlessly with customers proprietary systems and protocols.

- Training and Support: The cost of training customers personnel on how to effectively use the AI product for predictive maintenance. This includes creating training materials, conducting training sessions, and providing ongoing technical support to address any issues or questions that may arise during product usage.
- Maintenance and Upgrades: The cost of maintaining and updating the AI product over time. This includes monitoring the performance of the system, applying bug fixes and performance enhancements, and periodically retraining the models to ensure accuracy and relevance as new data becomes available.
- Operational Costs: Consider ongoing operational expenses such as energy consumption, software licenses, monitoring tools, and any additional personnel required to manage and operate the AI product within customers environment.
- Downtime and Risk Mitigation: Account for the potential costs associated with system downtime, including lost productivity, revenue, and customer dissatisfaction. Implementing backup systems, redundancy measures, and disaster recovery plans can help mitigate such risks.

TCO shall be transparently communicated to the customers during every sales phase.

CHAPTER 10 RISK ASSESMENT

When conducting a risk assessment for a new AI product for predictive maintenance in automotive testing, consider the following factors:

- Data Quality and Availability: Assess the quality, availability, and reliability of the data required for the AI product. Consider factors such as data accuracy, completeness, relevance, and potential biases. Inadequate or biased data can lead to incorrect predictions or unreliable maintenance recommendations.
- Model Performance and Accuracy: Evaluate the performance and accuracy of the AI model for predictive maintenance. Conduct thorough testing and validation to ensure the model provides reliable and precise predictions. Assess the model's performance metrics, such as precision, recall, and F1 score, to gauge its effectiveness.
- False Positives and False Negatives: Analyze the potential impact of false positives and false negatives generated by the AI product. False positives may result in unnecessary maintenance actions, leading to increased costs, while false negatives could lead to missed maintenance opportunities and potential safety risks. Determine acceptable thresholds and assess the consequences of misclassifications.
- Integration and Compatibility: Evaluate the compatibility and integration of the AI product with existing automotive testing systems and infrastructure. Assess potential technical challenges, system disruptions, or compatibility issues that may arise during implementation. Consider the feasibility of integrating the AI product into the existing workflow and processes.
- Regulatory Compliance: Ensure the AI product complies with relevant regulations, standards, and guidelines in the automotive industry. Consider data privacy, security, and compliance requirements, as well as any specific regulations related to predictive maintenance and AI

technologies in automotive testing. Address any potential legal or ethical concerns.

- Scalability and Performance: Assess the scalability and performance capabilities of the AI product. Consider the system's ability to handle large volumes of data, real-time processing requirements, and potential bottlenecks. Evaluate the computational resources, infrastructure, and scalability plans to accommodate increasing data volumes and user demands.
- User Training and Understanding: Determine the training needs and level of understanding required for users of the AI product. Ensure that users receive appropriate training to interpret and act upon the product's recommendations effectively. Consider the usability of the product's interface and any potential user experience challenges.
- Business Impact: Analyze the potential business impact of the AI product. Consider the cost savings, efficiency gains, and improved maintenance outcomes that the product can deliver. Evaluate the return on investment (ROI) and assess the product's alignment with the organization's strategic objectives and goals.
- Continuous Monitoring and Maintenance: Establish a plan for continuous monitoring and maintenance of the AI product. Consider the need for regular model retraining, data updates, and performance evaluation. Ensure that there are mechanisms in place to detect and address any degradation in the model's performance over time.

See Risk Map on Figure 25 Risk Matrix

Contingency plans to mitigate potential risks and address any issues that may arise during the sales, the implementation and operation of the AI product shall be developed for every big customer. Backup measures, alternative solutions, and recovery strategies shall be identified to minimize the impact of potential failures or disruptions.

CHAPTER 11 BUSINESS CONCLUSION/RECOMMENDATION

Scenario 1: “New product sales via partner company who has own developed sales channels” fits company strategical goal.

- It will allow to make organization transformation from IT outsourcing model to product development one.
- The company will get asset (product) which will generate new cash flow (see Figure 24 Scenario 1 Present Value of free Cash Flow) and will allow to achieve better antifragility.
- There is higher probability for success and shorter time to market working together with partner company.

BIBLIOGRAPHY

1. https://escholarship.org/uc/item/1f03f6hb
2. https://www.kdnuggets.com/2019/10/good-data-scientist-beginner-guide.html
3. https://towardsdatascience.com/how-to-implement-machine-learning-for-predictive-maintenance-4633cdbe4860
4. https://towardsdatascience.com/how-to-use-machine-learning-for-anomaly-detection-and-condition-monitoring-6742f82900d7
5. https://www.transparencymarketresearch.com/automotive-predictive-maintenance-market.html

List of Figures

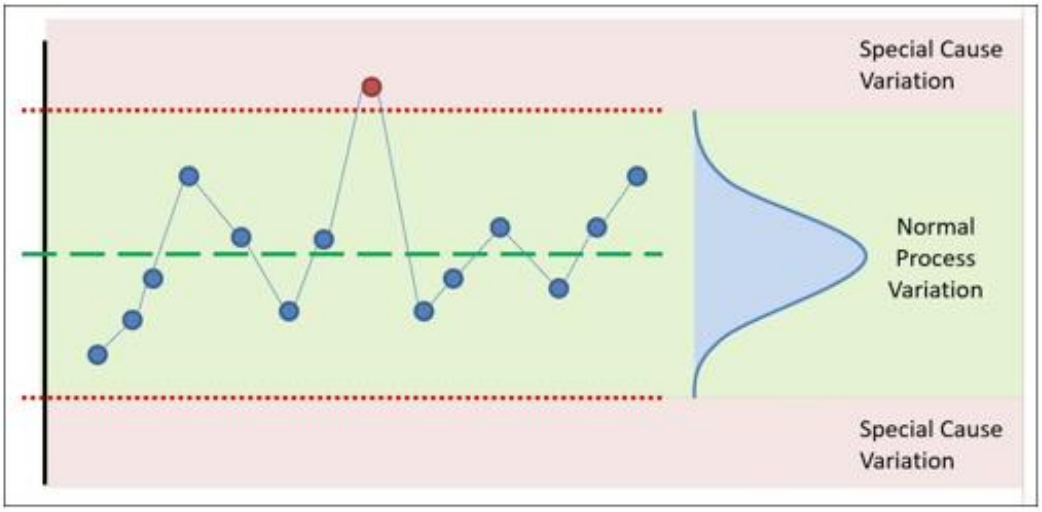

Figure 1 Statistical methods for anomaly detection.

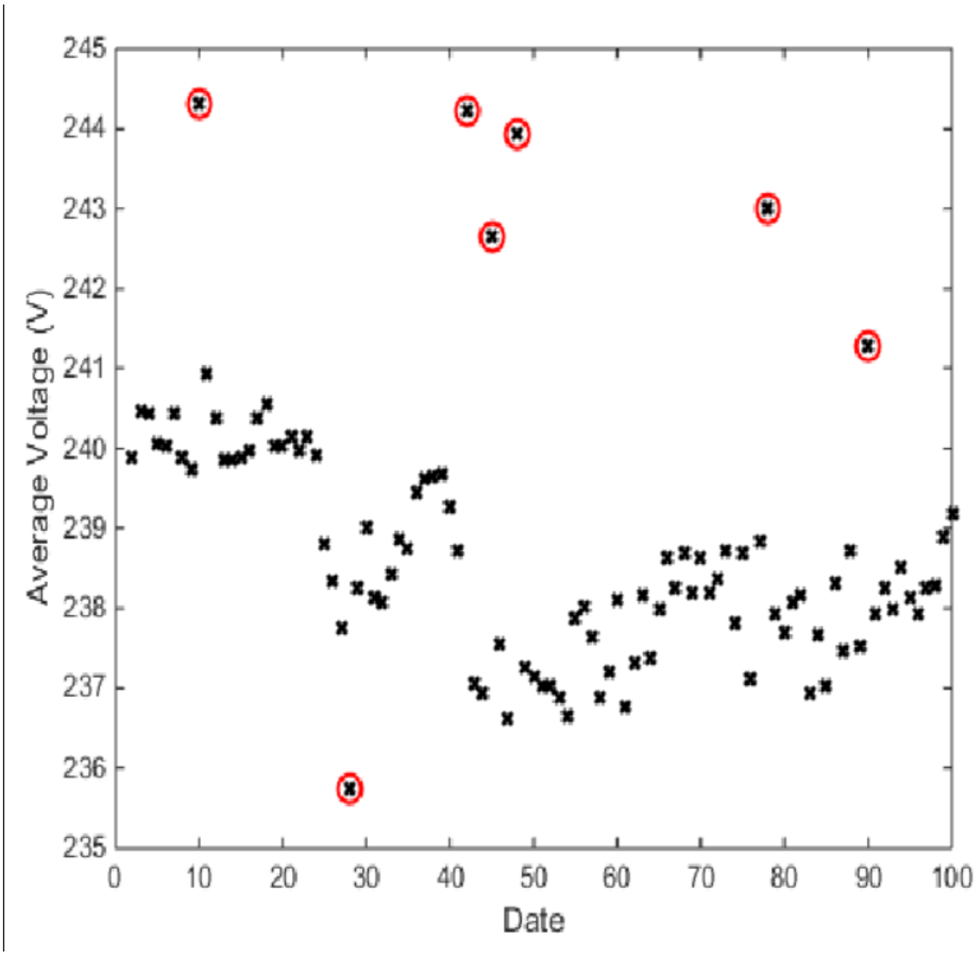

Figure 2 Examples of point anomalies

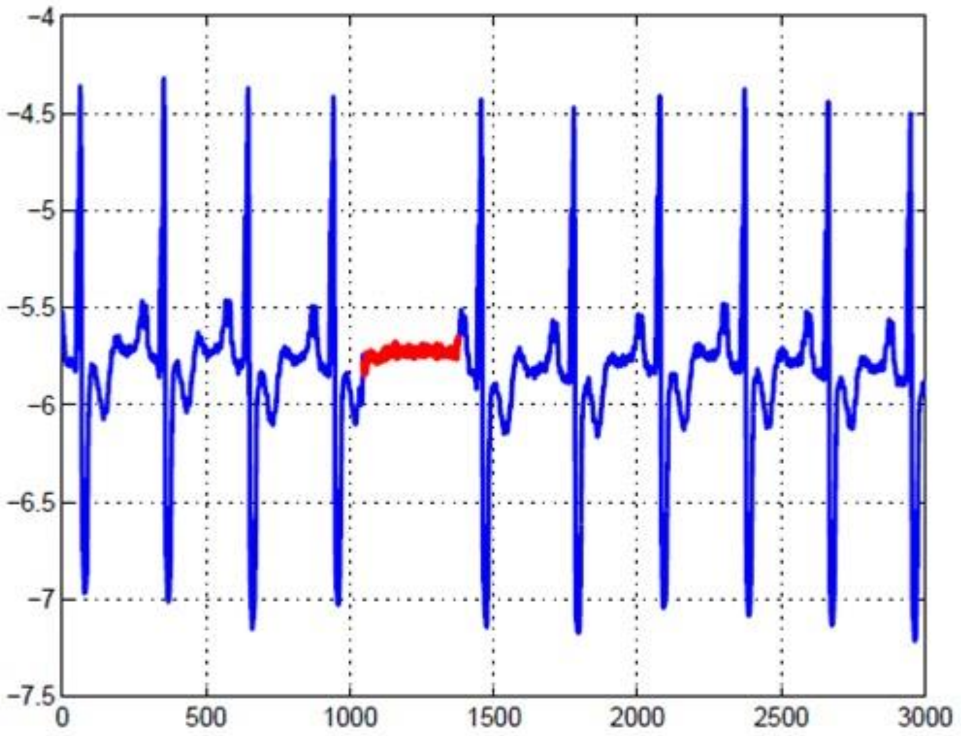

Figure 3 An irregular heart beat is an example of a collective anomaly.

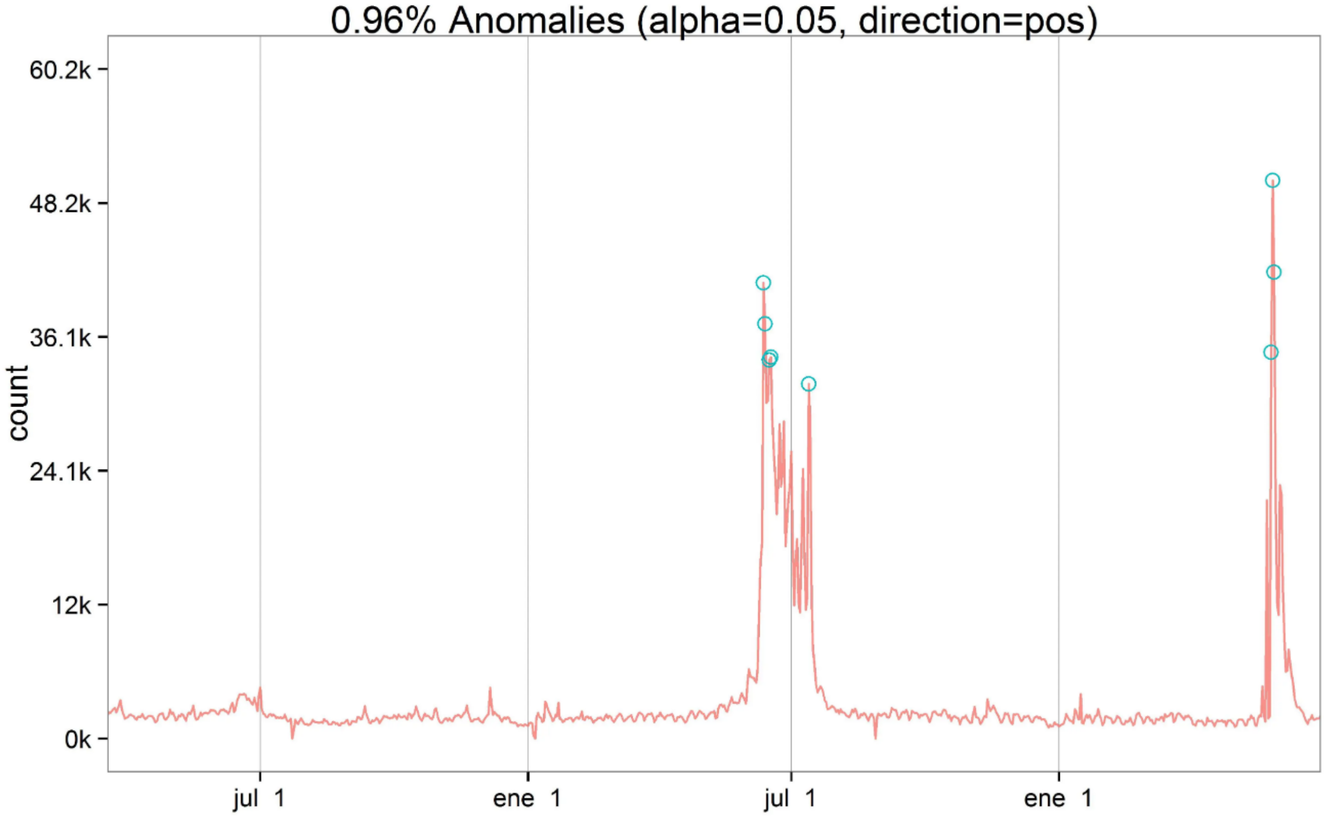

Figure 4 Example of contextual anomaly detection using the Twitter AnomalyDetection package in R

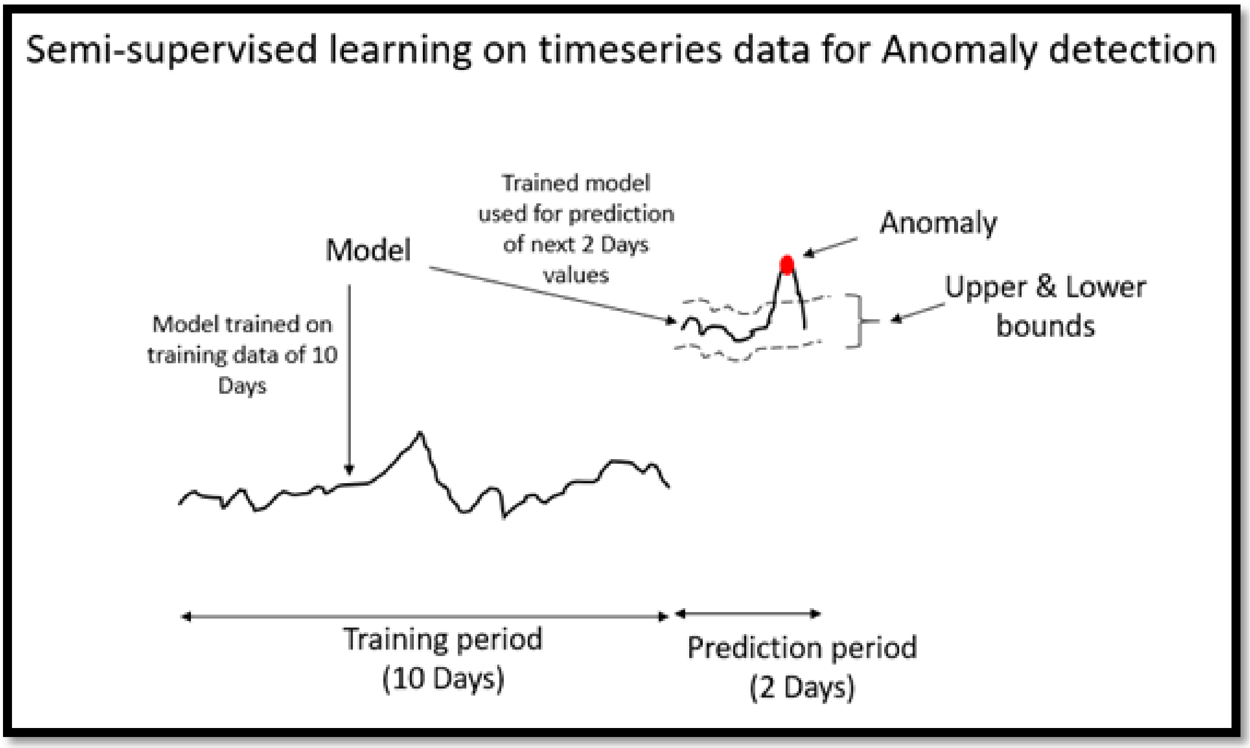

Figure 5 Example of semi-supervised learning on timeseries data for Anomaly detection

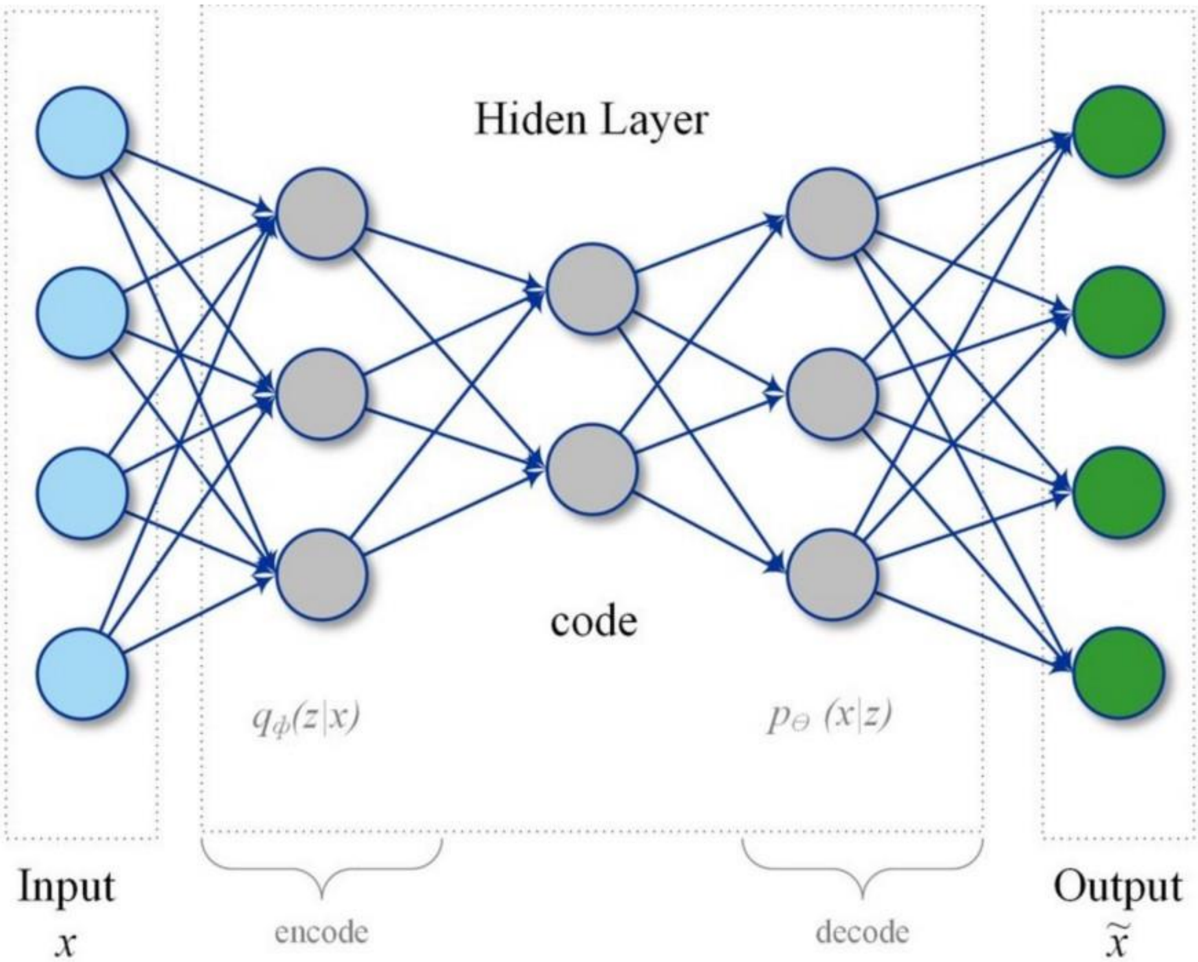

Figure 6 Autoencoder network

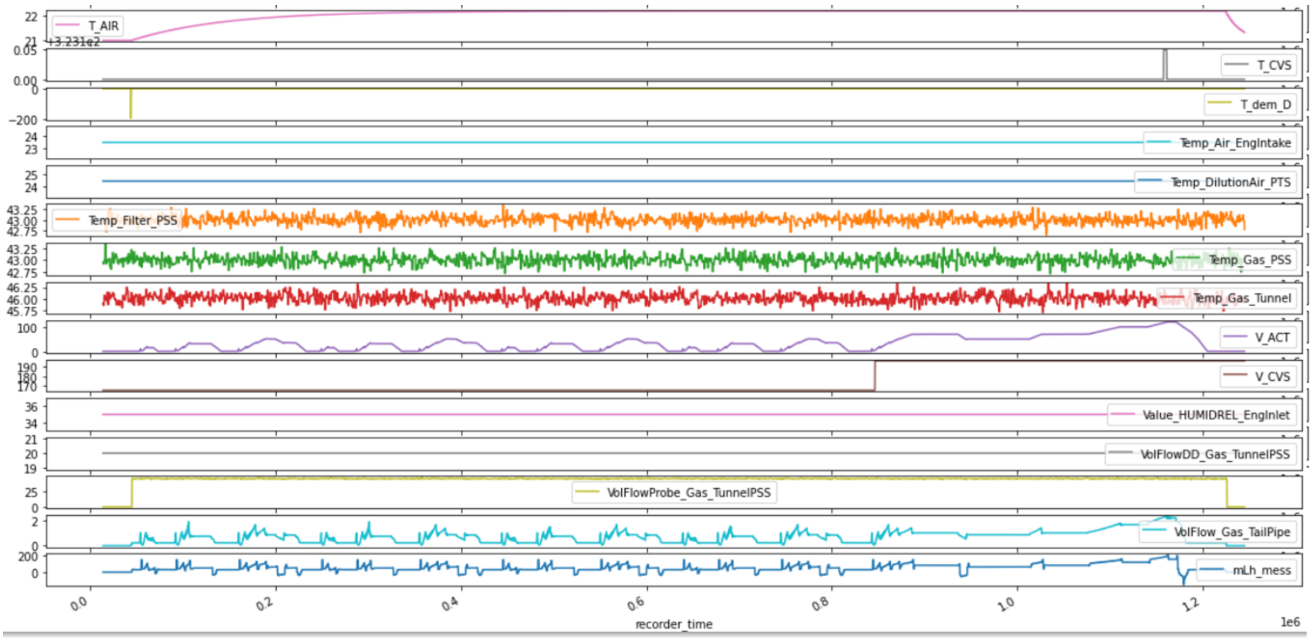

Figure 7 numeric continuous features of the dataset

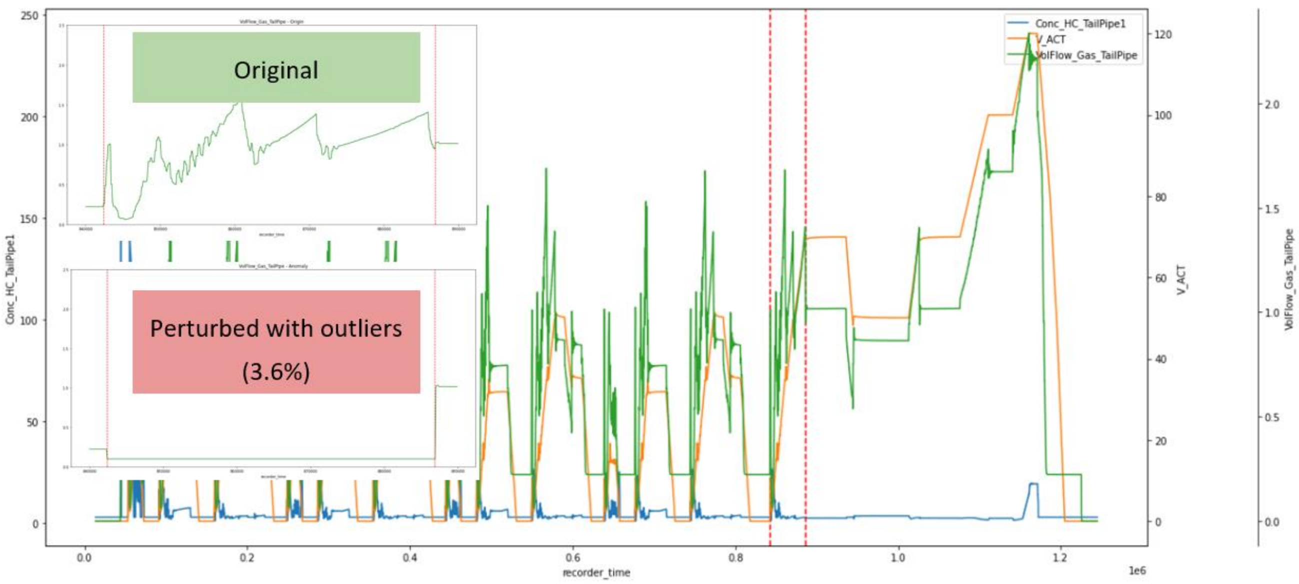

Figure 8 Dataset was perturbed with 3.6% outliers from total number

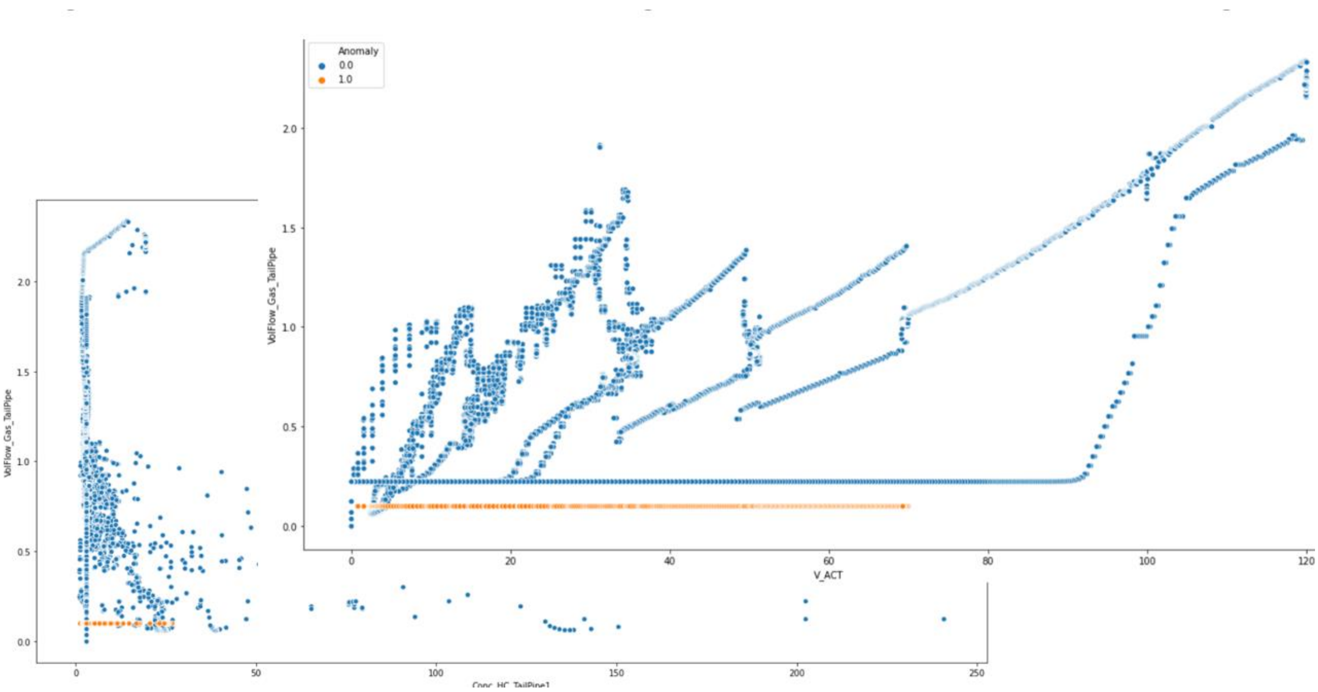

Figure 9 Exploratory data analysis shows labeled outliers in perturbed dataset

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_5 (Dense)              (None, 10)                40
_________________________________________________________________
dense_6 (Dense)              (None, 2)                 22
_________________________________________________________________
dense_7 (Dense)              (None, 10)                30
_________________________________________________________________
dense_8 (Dense)              (None, 3)                 33
=================================================================
Total params: 125
Trainable params: 125
Non-trainable params: 0
```

Figure 10 AutoEncoder neural network model

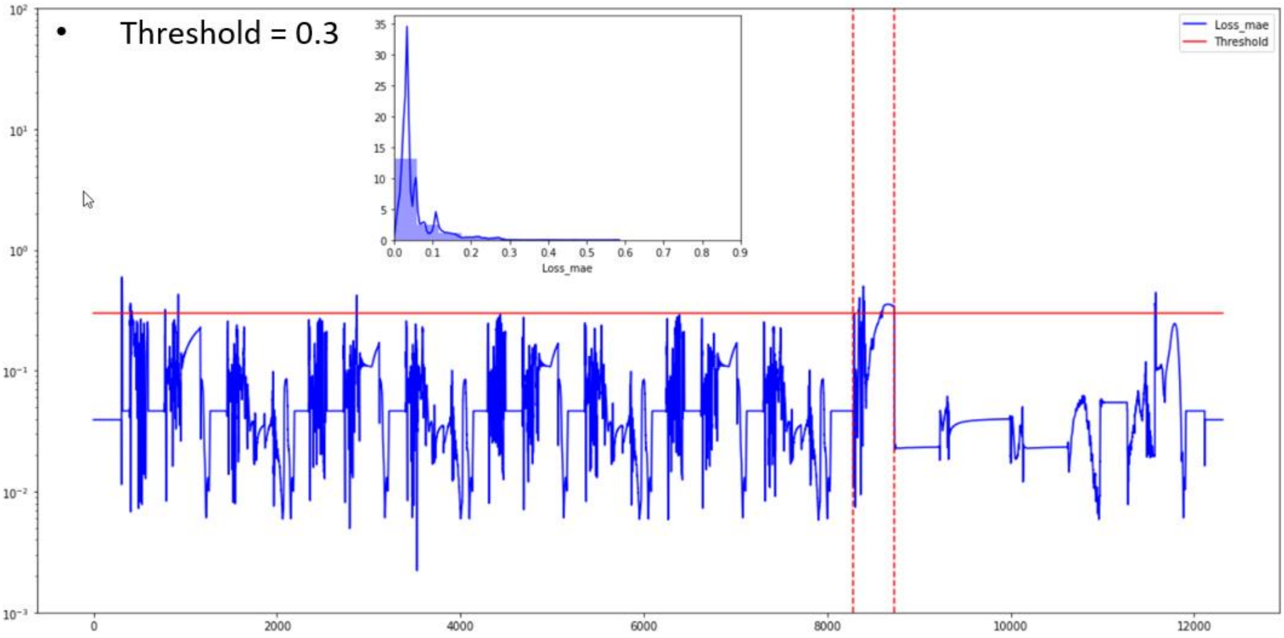

Figure 11 Threshold selection

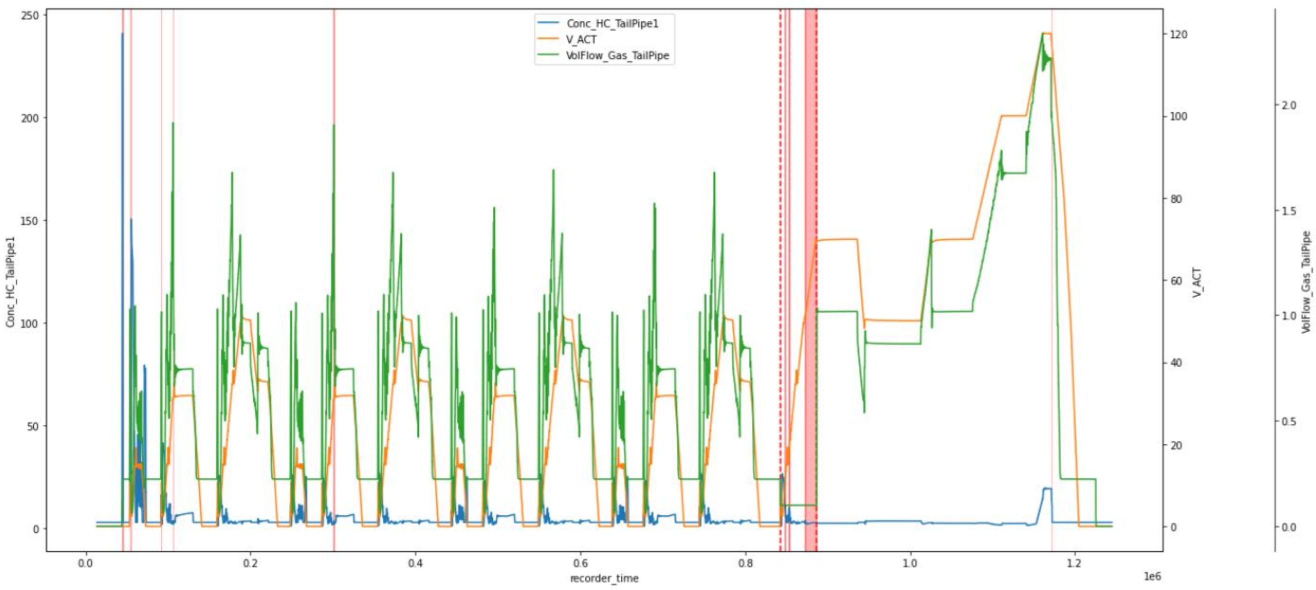

Figure 12 Outlier prediction

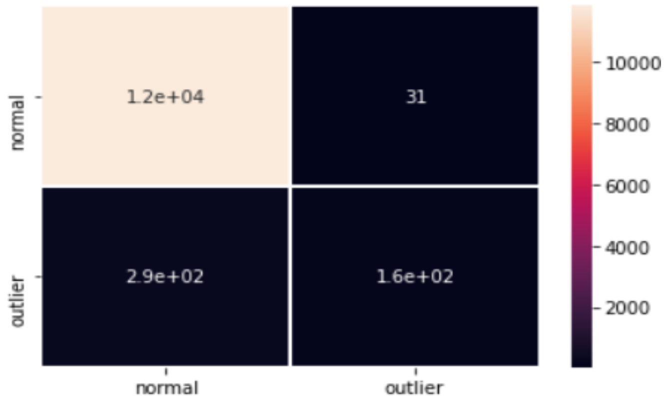

Figure 13 AutoEncoder accuracy

Strength	**Weaknesses**
• Available software (SW) development infrastructure and corresponding human capital • Proved successful track of working with foreign customers • Automotive/Emission testing expertise • COO MBAI • Availability of some financial resource for investment	• War in Ukraine • Dependency to one main customer with only outsoursing (rather outstaffing) orders with fixed (rather low) rate • Staff/ Tecnological stack aging: there are no web/cloud and data science stack etc • Less compatative for human capital acquisition/ retention in comparison with other SW companies which have more resources/assets • No marketing/sales expereince
Opportunities	**Threats**
• SW development and data science R&D demand growth • Automotive industry transformormation with focus on electrification, SW usage and efficiency • Increased focus on reducing unplanned maintenance downtime is boosting the automotive predictive maintenance market.	• Order reduction from current main customer and future potential customers because of political and economical factors (our customer is on transformation process because of electrification disruption, industry stagnation etc) • Ukraine tax reform which reduces SW development business marginality

Figure 14 Company internal SWOT analysis

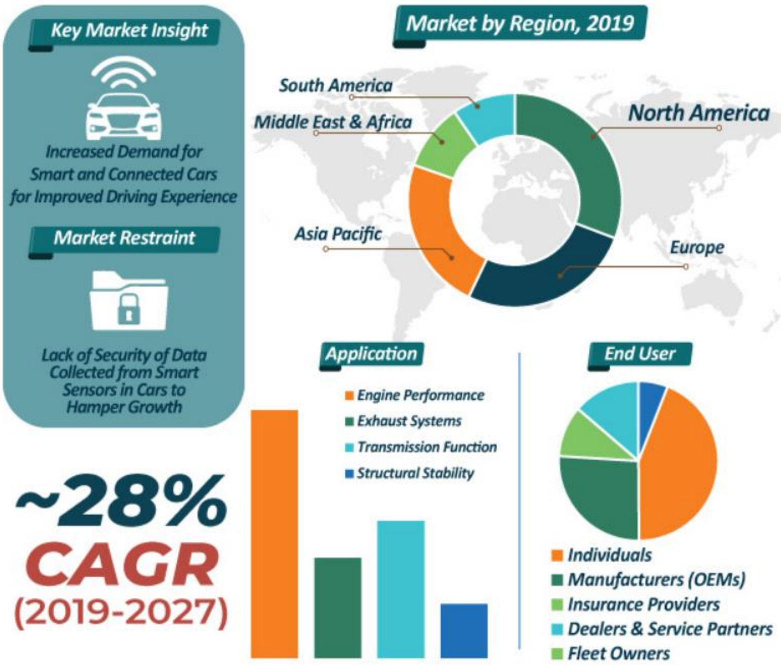

Figure 15 Market Micro-environmental audit for automotive predictive maintenance

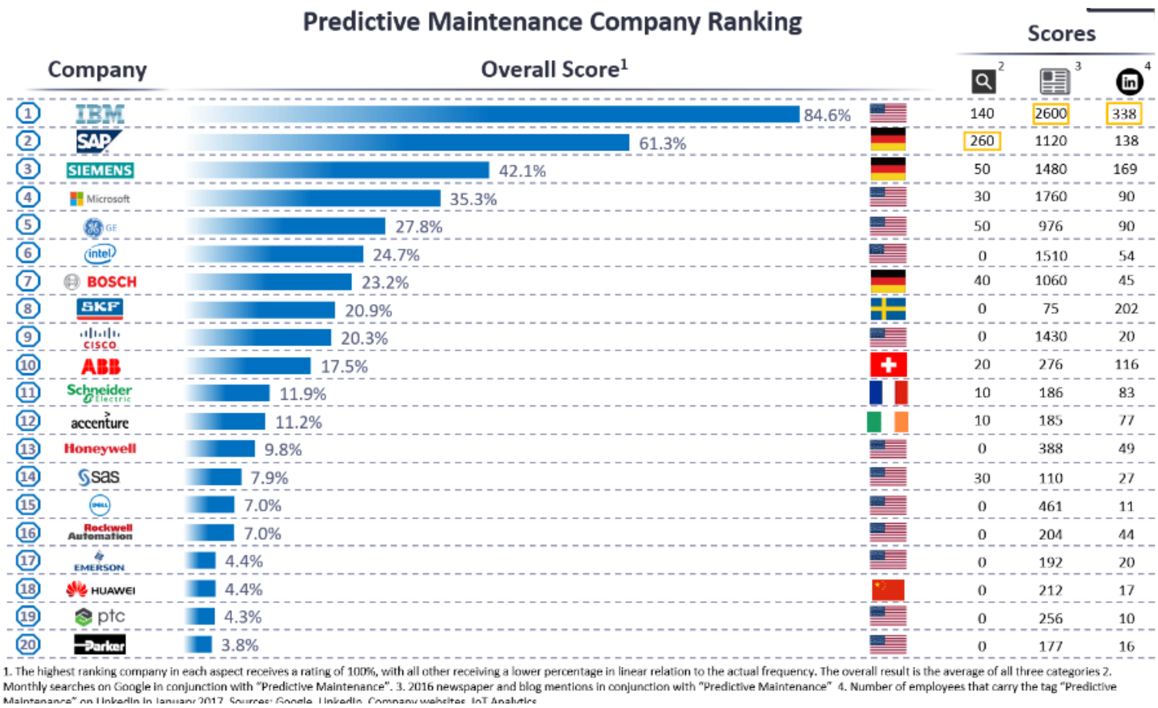

Figure 16 Predictive maintenance company ranking

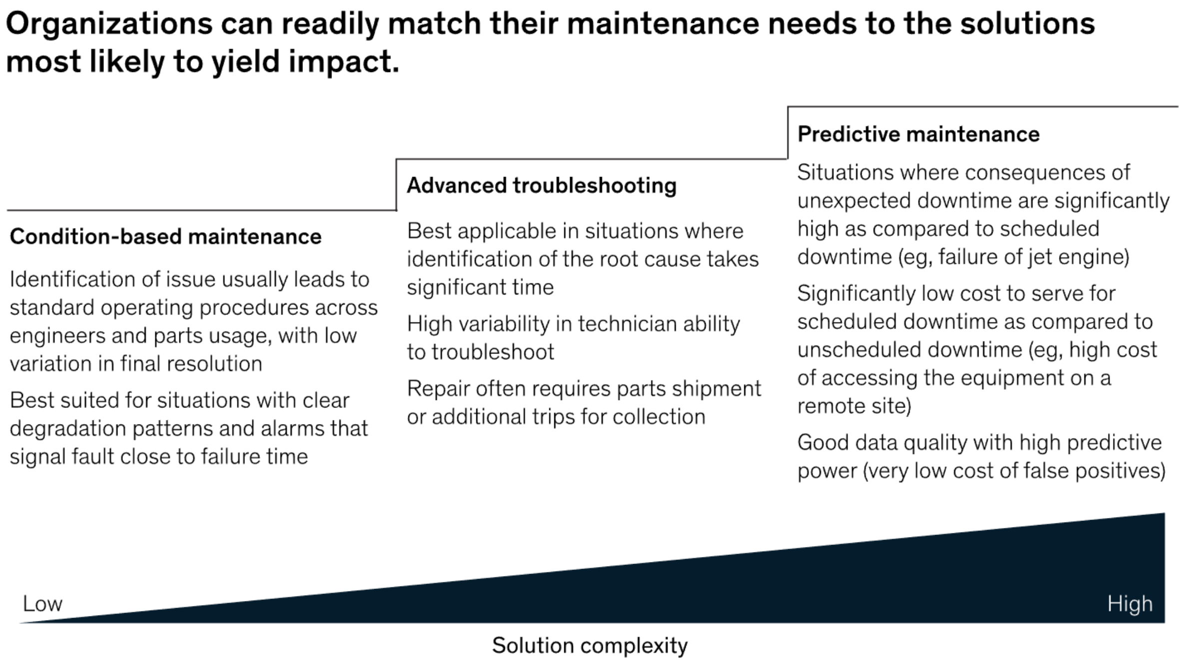

Figure 17 Alternative maintenance options

		Units	Abs.	Rel.
	Ours financing required	USD	**190.000**	
	Development	months	12	
	Start-up period	months	18	
	Expenses, including:	USD	185.040	100%
1	AI Data Scientiests and Software engineers - Development	USD	96.000	52%
2	AI Data Scientiests and Software engineers -Maintenance for 3 years	USD	43.200	23%
3	G&A expenses - Development	USD	19.200	10%
4	G&A expenses -- Maintenance for 3 years	USD	8.640	5%
5	Computers, Hardware, Software	USD	18.000	10%
	Credit financing	%	0%	
	Equity financing	%	100%	
	Equity financing required	USD	190.000	
	Plus, the sum required to cover % interest expense in start-up period	USD	0	
	Credit financing required	USD	0	
	Partner's financing required	USD	**135.000**	
1	Marketing and sales expenses for 3 years		135.000	
	Additional assumptions:			
	Average subscription fee per year	USD	60.000	
	Number of sold subscriptions in 1st operating year	-	1	
	Number of sold subscriptions in 2nd operating year	-	3	
	Number of sold subscriptions in 3d operating year	-	5	
	Ourr Revenue share	%	58%	
	Partner's Revenue share	%	42%	
	Tax Rate	%	18,0%	

Figure 18 Scenario 1 Investment highlights

	Units	Pre-Revenue Period 0 2024	REVENUE OPERATIONS 2025	2026	2027
Profit and Loss Statement (P&L)					
Net Revenue without VAT	USD	0	35.077	105.231	175.385
- Cost of Goods Sold (Cost of Sales)	USD	96.000	14.400	14.400	14.400
Gross Profit	USD	-96.000	20.677	90.831	160.985
Gross profit margin =	%	0	59%	86%	92%
- Selling General and Administrative Expenses	USD	19.200	2.880	2.880	2.880
Operating Profit (< = > EBITDA)	USD	-115.200	17.797	87.951	158.105
Operating profit margin =	%	0	51%	84%	90%
- Depreciation and Amortization	USD	0	3.600	3.600	3.600
EBIT	USD	-115.200	14.197	84.351	154.505
EBIT profit margin =		0	40%	80%	88%
- Interest Expenses		0	0	0	0
EBT (Taxable amount)	USD	-115.200	14.197	84.351	154.505
EBT profit margin =		0	40%	80%	88%
- Taxes	USD	0	2.555	15.183	27.811
Net Income	USD	-115.200	11.641	69.168	126.694
Net income margin		0	33%	66%	72%
Income distribution					
Net Income	USD	-115.200	11.641	69.168	126.694
DIV payout	%	0%	0%	0%	0%
Dividends	USD	0	0	0	0
Retained earnings	USD	-115.200	11.641	69.168	126.694
Cumulative retained earnings	USD	-115.200	-103.559	-34.391	92.303

Figure 19 Scenario 1 Profit and Loss statement, Income Distribution

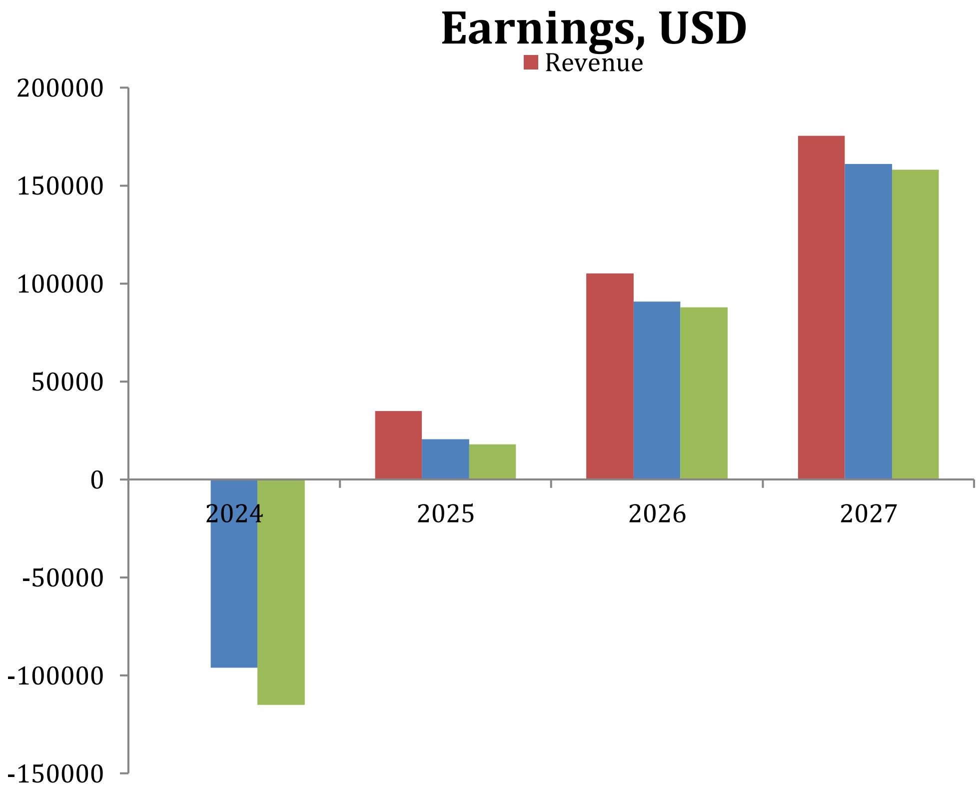

Figure 20 Scenario 1 Earnings graph

Balance Sheet					
ASSETS					
Current assets					
Cash & Cash Equivalents	USD	56.800	72.041	144.809	275.103
Accounts Receivable	USD	0	0	0	0
Inventory	USD	0	0	0	0
Prepaid expenses	USD	0	0	0	0
Total current assets	USD	56.800	72.041	144.809	275.103
Fixed assets (non-current assets)					
Gross PPE	USD	18.000	18.000	18.000	18.000
- cumulative depreciation	USD	0	3.600	7.200	10.800
Net Fixed Assets	USD	18.000	14.400	10.800	7.200
TOTAL ASSETS	USD	74.800	86.441	155.609	282.303
LIABILITIES AND SHAREHOLDERS' EQUITY					
Accounts Payable	USD	0	0	0	0
Short-term liabilities	USD	0	0	0	0
Current liabilities	USD	0	0	0	0
Credits	USD	0	0	0,0	0,0
Bonds	USD	0			
Total Liabilities	USD	0	0	0	0
Contributed capital	USD	190.000	190.000	190.000	190.000
Retained earnings (accumulated capital)	USD	-115.200	-103.559	-34.391	92.303
Shareholders' equity	USD	74.800	86.441	155.609	282.303
LIABILITIES AND EQUITY	USD	74.800	86.441	155.609	282.303
Balance check line	USD	0	0	0	0
Balance check line		TRUE	TRUE	TRUE	TRUE

Figure 21 Scenario 1 Balance Sheet

Cashflow Statement					
Net Income	USD	-115.200	11.641	69.168	126.694
Depreciation	USD	0	3.600	3.600	3.600
+ or - changes in AR	USD	0	0	0	0
+ or - changes in AP	USD	0	0	0	0
Cashflow from Operating Activities	USD	-115.200	15.241	72.768	130.294
Capital Investments (CapEx)	USD	-18.000	0	0	0
Return (dividends) from investments made	USD				
Cashflow from Investing activities	USD	-18.000	0	0	0
Credits from commercial banks	USD	0	0	0	0
Repayment of credits to commercial banks	USD	0	0	0	0
Proceeds from IPO	USD	0	0	0	0
Equity financing from the founders/investors	USD	190.000	0	0	0
Dividends to the shareholders	USD	0	0	0	0
Cashflow from Financing Activities	USD	190.000	0	0	0
Total Cashflow	USD	56.800	15.241	72.768	130.294
Cash at the beginning of the period	USD	0	56.800	72.041	144.809
Cash at the end of the period	USD	56.800	72.041	144.809	275.103

Figure 22 Scenario 1 Cashflow statement

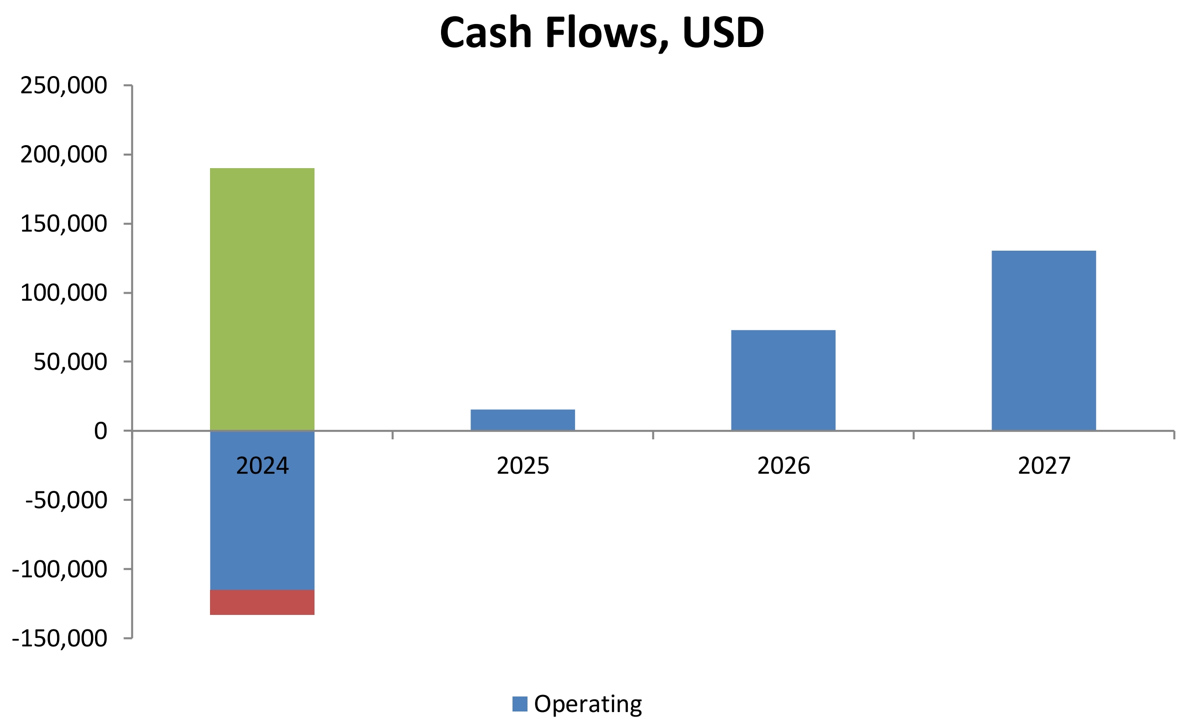

Figure 23 Scenario 1 Cashflow graph

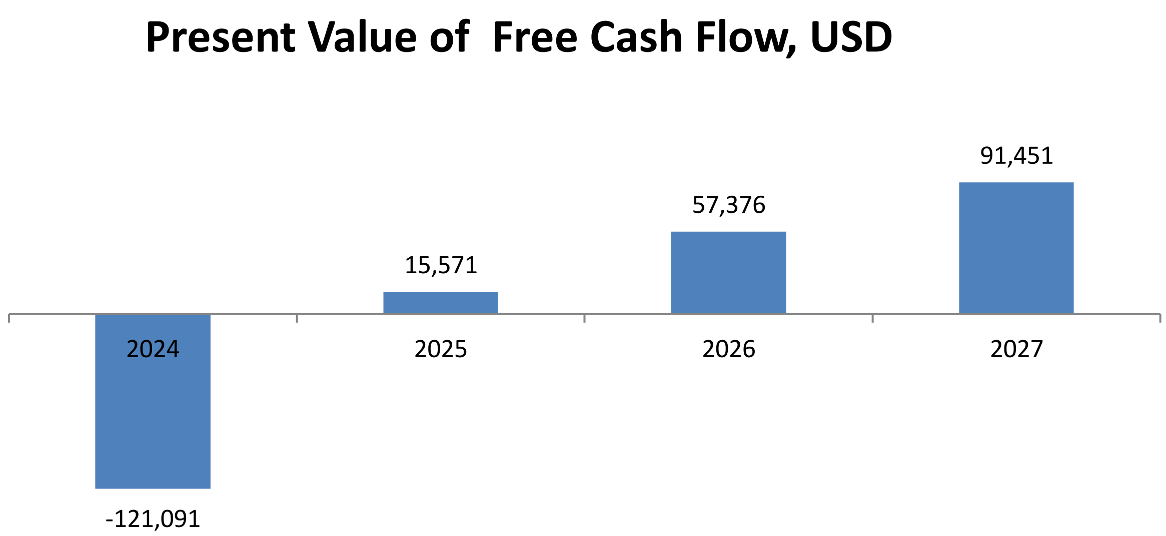

Figure 24 Scenario 1 Present Value of free Cash Flow

KSE | Business Education

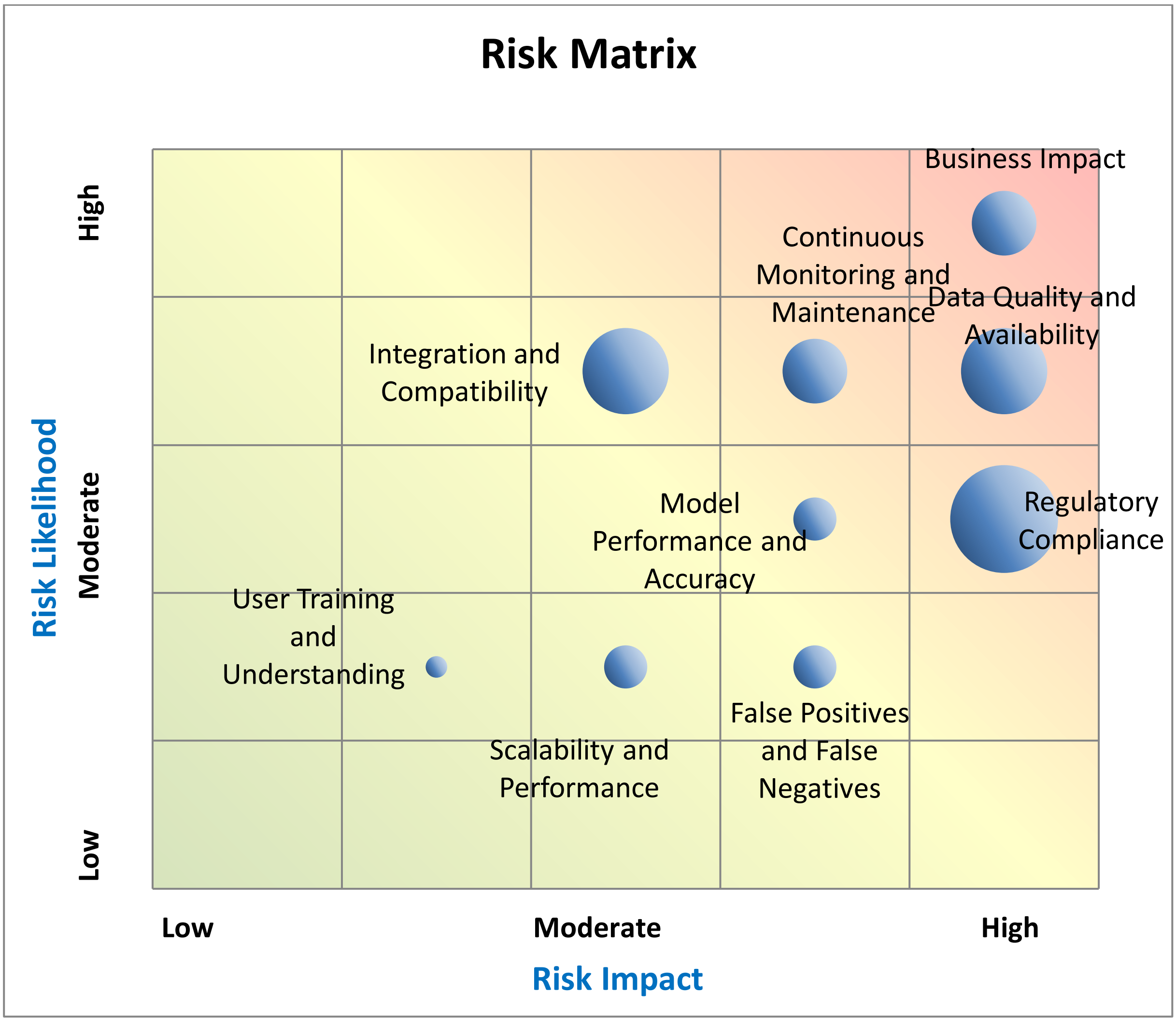

Figure 25 Risk Matrix

Appendix Project Code

The code is written on Python within Jupyter Notebook and is attached to the document.

OlegMoroka_CapstoneProject.ipynb