Enhancing quality assurance by automating manual work with AI and Machine Learning

Using AI and machine learning to simplify, modernize, streamline processes, enhance quality assurance, and reduce costs for the production line.

Industrial and Materials Science

Arvid Sundberg and Gavin Man

THE INSTITUTION FOR INDUSTRIAL AND MATERIAL SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2024

www.chalmers.se

BACHELOR'S THESIS 2024

Enhancing quality assurance by automating manual work with AI and Machine Learning

Gavin Man
Arvid Sundberg

Department of Industrial and Materials Science

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2024

Enhancing quality assurance by automating manual work with AI and Machine Learning

GAVIN MAN
ARVID SUNDBERG

Supervisors:
Robert Krickic, Head of Manufacturing Engineering VM&C

Examiner:
Peter Hammersberg, Department of Industrial and Materials Science, Chalmers University of Technology

BACHELOR'S Thesis 2024
Department of Industrial and Materials Science and Engineering
Division of Mechanical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31772 1000

Cover picture: The Volvo Cars manufacturing plant in Torslanda, Gothenburg.

Type set in LaTeX
Gothenburg, Sweden 2024

Enhancing quality assurance by automating manual work with AI and Machine Learning
GAVIN MAN
ARVID SUNDBERG
Department of Industrial and Materials Science and Engineering
Chalmers University of Technology

Abstract

The rapid advancement in technology, drives towards modernization and automation which gives opportunities for all engineers. This thesis researches the potential of using Machine Learning (ML) and Artificial Intelligence (AI) to automate the manual work in a production line. This is a collaboration between Chalmers University of Technology and Volvo Cars. The focus has been on the borging process at Volvo Cars, a critical quality assurance that involves manually inspecting and verifying the precision of screws in car assembly. The primary objective is to identify innovative cost-saving strategies and increase efficiency by integrating AI and ML into the production line while ensuring safety and maintaining high-quality standards.

The automotive industry is continuously introducing new models, necessitating adjustments to the assembly processes to fit different parts and materials. Every new model and improved material goes through multiple testing and validation phases, which currently include a manual borging process. This manual process involves checking the screws using torque wrenches and visual inspections to ensure that they meet specified standards. Data from these checks is collected, stored, and compared against predefined values (PKI values).

This thesis dives into the possibilities of automating or semi-automating the borging process to reduce manual work, lower costs, and improve production efficiency. The research methodology includes literature studies, data analysis, interviews with industry experts, and site visits to Volvo Cars production facilities. These activities provided a detailed understanding of the current process.

Replacing manual work with AI in the production line will save the company labor costs, increase efficiency, and enhance safety for both cars and operators. This will result in more consistent outcomes by minimizing the human factor. Currently, multiple external workers are required to check all the screws on the cars in the production line. By replacing that with AI, the labor cost will be reduced massively.

Python is the chosen tool to develop the program and research due to its adaptability for machine learning tasks and relevance for training ML models. Various ML techniques, including supervised and unsupervised learning, were explored to predict the status of screw tightening.

The result of this research highlighted the benefits of using AI and ML in the assembly process. This thesis includes a final code that works on a smaller scale but is adjustable to all the different screws and stations in the production line. To maintain high safety when making predictions about the tightening process, a two-step verification method is implemented.

Keywords: Artificial intelligence, Machine Learning, Python, Algorithm, Cars, Production line, Assembly line, Quality assurance, Automating, Semi-automating and Manufacturing

Acknowledgements

We would like to extend our heartfelt gratitude to the entire team at Volvo Cars, especially the pre-trim division, for their exceptional help, assistance, and guidance throughout our project. Your support has been instrumental in the successful completion of our work.

In particular, we would like to express our deepest appreciation to Robert Krickic for granting us this invaluable opportunity. His encouragement allowed us to deepen our understanding of artificial intelligence and machine learning, providing us with profound insights into the future of production. This opportunity to test and apply our knowledge in a real-world environment has been an enriching and invaluable experience.

Additionally, we owe a special debt of gratitude to Peter Hammersberg, whose unwavering assistance and support have been crucial during this thesis work. His expertise and willingness to help have greatly contributed to our project's success.

Thank you once again to everyone at Volvo Cars for making this experience both educational and memorable.

Arvid Sundberg

Gavin Man

Glossary

AI	*Artificial Intelligence* The simulation of human intelligence processes by machines, including learning, reasoning, and self-correction.
ML	*Machine Learning* A subset of artificial intelligence that focuses on developing algorithms that enable computers to learn from data and improve performance over time.
Algorithm	A set of instructions or rules followed by a computer to solve a problem or perform a task.
Semi-supervised learning	A machine learning paradigm that combines elements of supervised and unsupervised learning, typically using a small amount of labeled data and a large amount of unlabeled data.
Reinforcement learning	A type of machine learning where an agent learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties.
Online learning	Machine learning that updates and adapts in real-time as new data becomes available.
Neural Network	A computational model inspired by the structure and functioning of the human brain, used in machine learning and artificial intelligence for pattern recognition and data analysis.
Clustering	A method of unsupervised learning that involves grouping similar data points together.
Regression	A statistical method used to predict the relationship between variables, often used for forecasting and trend analysis.
AdaBoost	Short for Adaptive Boosting, is an ensemble learning algorithm that combines multiple weak classifiers to form a strong classifier by adjusting the weights of misclassified instances in each iteration.
Gradient Boosting	Gradient Boosting is an ensemble learning technique that builds models sequentially, each new model correcting errors made by the previous ones, using gradient descent to minimize the loss function.
SVM	*Support Vector Machine* is a supervised learning algorithm that finds the optimal hyperplane to separate different classes in the feature space with the maximum margin.

MLP	*Multi-Layer Perceptron* is a type of neural network consisting of multiple layers of neurons, including at least one hidden layer, used for complex pattern recognition and classification tasks.
1D-CNN	One-Dimensional Convolutional Neural Network is a type of neural network that applies convolutional filters along one dimension, commonly used for analyzing sequential data like time series or text.
Naïve Bayes	Naïve Bayes is a probabilistic classifier based on Bayes' theorem, assuming independence between features, often used for text classification and spam detection.
Dataset	A collection of data points used for analysis, often organized in a structured format.
Training data	Used to build and optimize a machine-learning model, and train it on a percentage of the dataset.
Test data	Test data is used to evaluate an algorithm's performance on unseen data based on training data.
ID	Identification number, used to identify a specific screw in a specific car.
PKI/PII	*Produkt kontroll instruktioner* or *Process and Inspection Instruction*, used in same context, the English and the Swedish version
Boxplot	A graphical representation of the distribution of a dataset, showing the median, quartiles, and outliers.
Python	A widely used high-level programming language known for its simplicity and readability, commonly used in data analysis, machine learning, and web development.
2-step verification	A security process that requires users to provide two different authentication factors to gain access to an account or system.
Torque wrench	A tool used to apply a specific amount of torque to a fastener, ensuring proper tightening.
PowerBI	A business analytics service by Microsoft that provides interactive visualizations and business intelligence capabilities.
OK	A Correct tightening
NOK	A Incorrect tightening

Contents

List of Figures

List of Tables

Introduction 1

Constant technological growth, the pursuit of modern production facilities, and automation of the car's production are increasing. Due to new challenges and innovation, it has become more difficult for engineers to adjust their plants to the new creations. To automate or semi-automate a part of a plant in the process is important for efficiency but also to ensure a safe production environment. Therefore, the automation needs to be added correctly, so it benefits the operator at the same time as it's getting more efficient. Due to the fast development in technology, testing occurs in test labs and different simulations to not directly affect the production line and to ensure that everything works. The benefit is that the deployment to the real environment will go faster and with fewer errors.

This thesis examines the challenges and potential benefits of rapidly integrating machine learning and AI into engineering, focusing on their global impact on modern production plants. Volvo Cars provided us the opportunity to research the possibility of using ML and AI in today's plant and how big of an impact it can have on many fields. Researching cost and safety benefits, as well as starting the discussion regarding how the future of production will look and fully rely on AI while still maintaining safety standards is important for the technology development in the production line.

Volvo Cars has several different models and more that are on their way to production shortly, which means that there are multiple ways to assemble the different parts. Different parts need different materials and ways to be assembled. When new models arrive or a better material or tool is found, changes are necessary for the process to ensure still the high safety measures that Volvo Cars holds as a company. When a new type of screw is introduced into the process multiple testing processes are being done, both simulations and physicals test. One of the tests is a process named "Borging" where an of screws in the cars are checked afterward to ensure the tools and screws are working as they should. This is done manually and in multiple steps to ensure that the cars delivered to customers are safe and well-built. All the data from every tightening is then collected and stored in a database. To know what values of the tightening are correct, they compare the values with a PKI that specifies the screw and process.

This thesis explores the different ways to automate or semi-automate the manual process while maintaining the safety of all the involved operators. The benefits of replacing the manual work include a reduced cost. Additionally, it is also more time-efficient, which will make it possible to produce more cars. Also, replacing manual work with AI or machine learning will ensure that both the company and the automotive industry stay at the forefront of technological advancement for the future.

1.1 Background

As the original project title suggests, "*Remove manual borging*" a deeper understanding of the borging process and the value of solving this issue was necessary, to be able to proceed further into exploring the potential benefits of AI and ML. Can AI and ML be applied to this case, and if so, how?

As mentioned before, Volvo Cars frequently produces new models and has many parts that need to be assembled with different screws and using different techniques. As new models are introduced, the screws need to be adjusted or changed for them to fit the parts, while also considering the legislation of safety demands and the properties of the screws, such as their endurance and the force that they can withstand. Today many tests are being done manually by humans which can affect the quality, since humans can be affected by the environment which can lead to faulty checks. By automating and adding a second layer of security to the process it's possible to increase safety.

So what is borging?

Borging is a crucial step for the quality assurance of the vehicles that are leaving the factories. This process ensures the precision of tightened screws, thereby guaranteeing the safety and reliability of every car produced. Currently, this task requires manual inspection by external operators who carefully examine all screws across a specified number of cars, using a torque wrench and visual inspection techniques.

1.2 Aim

The project involves exploring new innovative cost-saving strategies within the operational structure of Volvo Cars Torslanda's production line. Mainly to find different aspects of how manual borging can be removed or semi-automated by ML and AI. All findings of strategy for how to achieve the implementation of an ML system are beneficial. The project goal is to be able to replace the manual borging with a safe system that can operate with the datasets that are available today. The objective is to understand how we can efficiently work with AI and ML and find a long-term solution. By replacing manual work with ML and AI multiple operators in every station can be replaced with 1 supervisor with experience and knowledge of the program.

- Develop a method that can predict the tightening status.
- Understand the workflow of the operation in the assembly line.
- Compare different solutions for removing manual borging.
- Finding the optimal dataset and amount of data to use, finding a relationship in the dataset.
- Create a down-scaled version of what could be used in the future of production.

1.3 Research questions

The objective of the project is to answer the following questions:

- How will the production line look with AI and ML?
- Exploring the current readiness for introducing AI and ML, to replace manual borging
- Will and is it possible to reduce the workload and cost for a production line?
- Which type of machine learning is most optimal for the following case?
- Online or offline learning?
- What are the potential benefits?

To be able to answer all these questions research and tests will be done. Gaining knowledge from experts to site visits and programming will be involved. The following steps will be explained in Chapter 2 Methodology.

1.3.1 The value

Solving this case and being able to replace the manual workload in the production line is important. Although everything is working today, there are issues with the process, the operator's goal is to fasten the screw, and as the machine gives approval they move on. But there are many loopholes and ways to trick the machine and let faulty screws through. The operators work on a time limit and can therefore only rely on the machine. Letting faulty screws out of the factory can affect the quality and the safety of the customers.

The outcome and the consequences of the findings are not only for the company but also for the customers and workers. The company will be able to save labor costs by replacing multiple operators with 1 supervisor. More efficient and safer in the production line, being able to go from checking every tightening with a pen and torque wrench to a program. Being able to remove human errors from the quality assurance. The company can focus on other innovations with cost savings and not lower the quality. Also, stay in the technology development and apply this to other plants and stations all over the world. Even if this is a start of using ML and AI in a car production line in Torslanda for a specific station, it is a start of utilizing ML and AI in a production line overall to get a safer and more reliable outcome. The benefits show that is worth investing in ML and AI in other stations and other types of production lines. Not only from a cost-saving perspective but also from a time-saving and safety perspective.

Worth mentioning is that, in this state, Volvo Cars executes the checks for the tightenings manually and has not started to automate it. Therefore the span and limitations are so big and can therefore not include every aspect Most of the assumptions are based on tests and research. But it's necessary to break down all the parts in this thesis before continuing the development in the future.

1.4 Future work and Limitations

The request from Volvo Cars was to analyze existing data and utilize the statistics to find parameters for quality assurance (target, angle, and RPM). Develop a warning system to alert when tightening deviates from recommended parameters. Use AI and machine learning to achieve and make software. Also, create a table for all CC1 screw joints and be able to see in PowerBI.

Some limitations were set at the beginning, to scale down the problem and to be able to provide a quality product rather than giving a big quantity without quality and not making exaggerated assurances. By analyzing the data and the problem, and comparing it to the time frame and available knowledge, we narrowed down the issue.

The focus was set on finding a way to utilize a type of machine-learning algorithm to analyze and be able to predict new tightening as if they were applied in the factory. The main assignments are to create an effective and appropriate database for the situation and to develop a program that can make safe predictions for the tightening process on a smaller scale.

Which type of tightening process is best to start analyzing? The chosen machine for the tightening process, located at 1.6-798 in the factory, is used for the belt-tightening process. This station, like many others, follows three sets of tightening rules. Figure 1.1 illustrates a correct and an incorrect graph based on imaginary data for better visual understanding.

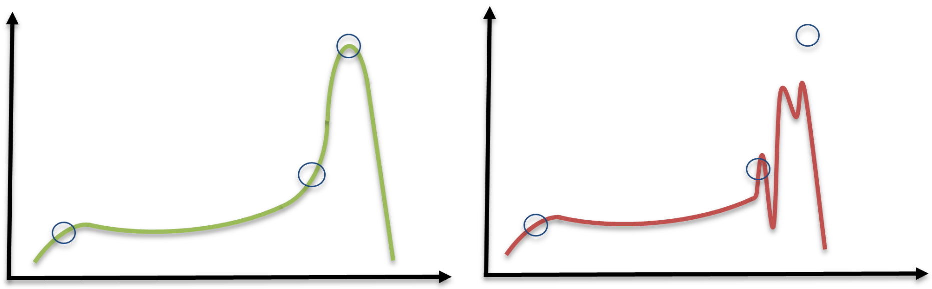

Figure 1.1: Example graph of a correct and incorrect tightening

The future of this project involves applying the program to various types of screws, each with its own set of rules and behaviors, which appear visually different on a graph. To achieve a final product within the provided timeframe and satisfy the company, the goal is to develop an AI or machine learning-based program capable of functioning in the given environment. The program must also be able to make predictions based on the selected dataset. A dataset analysis will be conducted, and future investigations of existing material will be analyzed and applied if suitable. Ideas that cannot be implemented within the timeframe will be presented to the company for future development. This project does not include code deployment and is intended only for a smaller scale.

1.5 Outline of the report

The purpose of the thesis is to research if it's possible to implement AI and machine learning on a production line. Furthermore, this project may provide potential solutions or ideas for further development and a better understanding of the processes. The purpose is also to provide a good working machine learning system that can operate with the type of dataset that is available. At the same time, by investigating the problem and different relationships with machine learning it provides research for academics and engineers. This research will be conducted in collaboration with Volvo Cars, focusing on their production line and the manual borging process.

1.6 Project provider introduction

Volvo Cars The word Volvo is Latin and stands for "I roll", Volvo Cars focuses on the safety of the cars. 1927 was the year of Volvo Cars' first production of cars in Gothenburg, Sweden, it was founded by Assar Gabrielsson and Gustaf Larsson. The objectives of Volvo Cars have always been to build quality and safe vehicles, the company stated according to Wikipedia. (2024) "The guiding principle behind everything we make is and must remain, safety". The first car left the factory at Hisingen, Gothenburg in 1927 and was called Volvo ÖV 4 also known as "Jacob". Volvo's increasing success in Sweden was due to the focus on safety and reliability but most importantly they had access to local steel that produced the car.

The company has always satisfied human needs and aims to make luxury and safe cars and hit its first millstone of 10000 cars in May 1932 and made cars that were affordable to most people. Since then, they continued to produce different car models for the market. 1944 Volvo was unveiled to become an international company, and "The Little Volvo" PV 444 was created and later on became a worldwide known car. in 1959 Bohlin that was an engineer at Volvo produced the three-point seat belt that was to prevent the body from being thrown forward. Volvo cars let the rest of the world use the innovation without any payment or claims to ensure that everybody is safe on the road. (Volvo Cars, n.d.-a)

Moving forward to 2013 the first manufacturing plant was completed outside of Europe in China, in the city of Chengdu, and was mainly producing S60L and XC60. In 2018 they continued to grow and made more manufacturing plants around the world one in Charleston, south California. Volvo Cars then became a truly global car manufacturer. (Our Heritage — Volvo Cars, 2013)

Volvo Cars aim is to continue to be a leading automotive industry in safety technology, electrification, and autonomous drive. They are still focusing on quality, safety, and luxury. The company headquarters remain in Volvo Cars' hometown Gothenburg. But have manufacturing plants in Belgium, the USA, and China and assembly plants in Malaysia and India. In figure 1.2, it's a look inside the production line. (Volvo Cars, n.d.)

Figure 1.2: Volvo Cars manufacturing plant in Torslanda, Sweden

2 Methodology

The project comprises several parts therefore the method is divided into different parts, the initial stage of the study was a literature study, research phase, interviews, and site visits, which is a common start to get a better understanding and different perspectives. The knowledge that was gained from the interviews was combined with the research information to get a better understanding of the whole project's aim and what the possibilities were. After the research phase, multiple decisions were going to be made, like what type of program and strategy to use to accomplish the goal. This chapter explains all decisions made and provides the reasoning behind each one.

2.1 Key performance indicators

Every minute in the factory multiple tightening is being executed on every station in the production line. This indicates that there is a large database to work with, to be able to start is important to know the amount of data that is going to be used.

The data can be downloaded in an Excel format with many different features and target values, therefore the dataset needs to be cleaned before using it. The dataset needs to be handpicked for it to work with the strategy that's going to be applied. The selection of which station and the type of screw is made by analyzing the existing data and by communicating with experts in respective areas. Testing and seeing the importance of every feature and target, one by one made it possible to pick the important features. All the targets are based on a PKI which is a standard to work from to control if the values are in target.

2.2 Programming

To work with machine learning and AI there needs to be a programming language that is adaptable and fits the aim of the study. Therefore, the chosen programming language was set to Python due to the adaptability to adapt to the problem and because if there is future development, Python is well-known for its machine learning and user-friendly environment.

2.3 Interviews

To start with the study and understand what vision the company has with removing the manual borging, multiple interviews and Q&A were requested. All the interviews within and outside of the company were conversational. All relevant persons revolving around the project were interviewed to get all information from every perspective. The questions were prepared beforehand and were specific to the person of expertise. The main people who were interviewed at Volvo Cars were a person with great knowledge of all the different screws and also someone handling the software that was being used for data collecting. Also, the person responsible for the assembly line, that controls the machinery. The project leader and an expert in data science were interviewed. Outside of Volvo Cars, an interview with one of the companies that provides Volvo Cars with tools was arranged to fully get an understanding of what was possible to achieve. Most of the questions were about the dataset that is available what every variable stands for and how the processes work. Questions about the vision and aim of the project were asked continuously to stay on the right path.

2.4 Site visit

To fully understand all the components in the assembly line and how the processes work from start to finish, there was a visit to the VCT assembly line. The observation in the VCT assembly line provided insight into the workflow in the pre-trim assembly line, the practical execution of the tightening process, and the operator's tasks from picking up screws to utilizing screwdrivers to apply on the cars. A clearer comprehension was gained regarding data collection methods and the construction of the workflow.

3 Literature studies

This part involves an exploration of the various subsets within machine learning to gain a deeper understanding of the most suitable approach. By researching different methods and techniques, a comprehensive analysis will be conducted to identify the most effective strategies for the specific case. Furthermore, an analysis of reports and articles will be researched to see how it's possible to implement machine learning and AI in similar contexts

3.1 Machine learning

Machine learning (ML) is a subset of artificial intelligence (AI) and is being used to do what a human can do. But faster and more efficiently, The ML collects data that is given by different systems or imputed but humans. With the help of different algorithms, it will process the data and learn from experience. The ML is going to find relationships and different patterns in the dataset that it's provided, to make predictions and help the human in the process. ML can help in different ways depending on the situation and the algorithms that are being applied.

The main purpose of different algorithms is to improve their performance to accomplish the task, over time the accuracy should increase and it should be more effective.

Taking a closer look at machine learning reveals its various branches, which include supervised learning, unsupervised learning, semi-supervised learning, and reinforced learning. This project will primarily focus on supervised learning, but all branches of machine learning will be carefully considered in the exploration, all the breaches as seen in Figure 3.3: Machine learning branches are connected to ML.

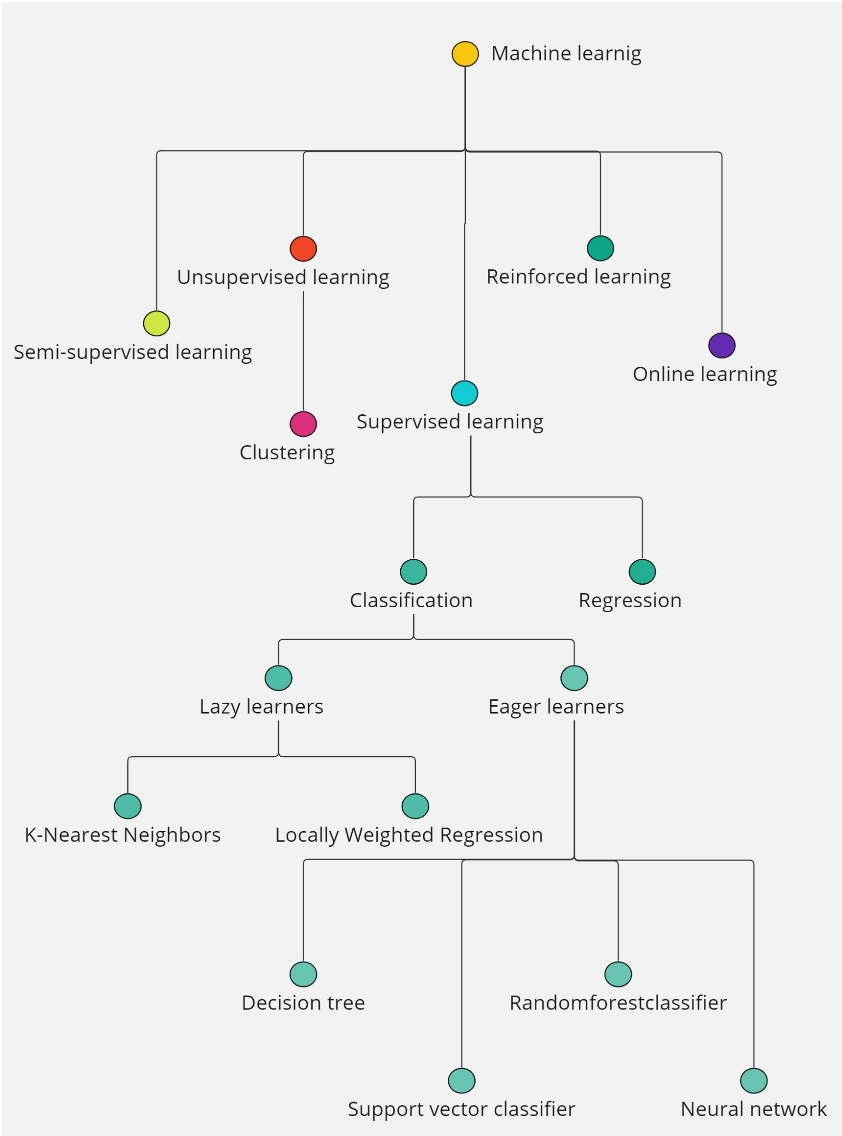

Figure 3.3: Machine learning branches

3.2 Supervised learning

Supervised learning uses an algorithm where the dataset has a label. The inputs are known as features and the outputs are known as targets. Supervised learning will try to identify the relationship between the input and the output. This is done by labeling the data, so the algorithm knows what's correct and incorrect. It's called supervised learning since it needs a supervisor who labels the dataset with a target. It also requires a large amount of data to get good accuracy and effective work, due to that it will cover a variety of possible cases, and the machine will learn and understand the pattern. Supervised learning is divided into two different groups, either classification or regression. (Bansal," Supervised and Unsupervised learning" 2024), (cloud.Google," What is Supervised Learning?", n.d.)

3.3 Classification

This is a subset of supervised learning and this method is used when the prediction can be labeled and classified for example either yes or no labels or 1 or 0 labels, so it needs to be a binary classification. It uses test and train data before getting used in an environment with unknown data. This means that it will train and test the accuracy with historical and label datasets before being performed with unseen data. The concept is to utilize known variables to make predictions. (Keita," Classification in Machine Learning: An Introduction", 2022). Classification is then divided into Lazy Learners and Eager Learners

3.4 Lazy learners

Lazy learners wait until they experience a new interaction, when it needs to make a prediction, it will then start to process the training data. This model will not build a model in the training set instead it will wait until it encounters a new interaction, this model is instance-based. But it will still store the memory that it has learned in the previous decision. The computational cost of a lazy learning algorithm can be expensive in the prediction phase. After all, it requires a long search process to find the right decision because it does not have a model and rules to base on. The most common algorithm in this category is K-NN. (What Is the Difference between Lazy and Eager Learning?, 2024)

Due to that Lazy learners start the process of the training data when they receive a new interaction it can be a benefit for problems with changing data so it can analyze the new data and compare it to the training set. If the dataset has a complex distribution or outliers, then this algorithm is best fitted because it will not influence the learning phase and not store the data that is odd. (Awan," What is Lazy Learning?", 2023)

But as mentioned before it can be slow when predicting data due to the working processes, it also needs a storage space to store the training data. Lazy learners can be sensitive to overfitting and noise because the predictions are often made by comparing them directly to the training set, so a large or small data set can impact the prediction negatively. (Awan," What is Lazy Learning?", 2023)

3.5 Eager learners

Eager learners are the opposite of Lazy learners, this ML will build a system that generalizes a model during the training phase to learn and find patterns and relationships before a new interaction appears. The approach is to feed the algorithm with a large dataset so it can train test and make a model, so it has learned before the prediction part begins. There are different algorithms in this subset, such as decision trees, neural networks, and support vectors. Eager learning is appropriate when the training set can be processed efficiently and can fit into memory. (Awan," What is Eager Learning?", 2023)

Eager learners are fast since already in the training set have built up a model and generalized the patterns, it's not sensitive to noise since it already has made global patterns for the predictions. But on the other hand, the training time takes longer, especially with large datasets. Eager learning can be sensitive to changing data and overfitting because it will try to generalize the data. (Awan," What is Eager Learning?", 2023)

3.6 Regression

Regression is the subset in supervised learning and it's very similar to Classification and the main difference is that regression predicts the values. Regression works with input/output data to predict unseen data. In the same way as classification, regression also tries to find relationships between input and output. In this subset the output is not binary instead it needs to be real or continuous value, examples are salary or other similar outputs. Regression techniques are from the formula as seen blow and are applied and modified in different ways to fit a certain case. The main types of regression techniques are Linear Regression, Polynomial Regression, and other regression. (Gupta," Linear Regression in Machine learning", 2024)

$$y = mx + c \tag{3.1}$$

3.7 Unsupervised learning

Unsupervised learning is different from supervised learning because it does not require a supervisor in the processes. It uses self-learning algorithms and techniques, in comparison to supervised learning it does not require labels or prior training. Instead, it works by inputting raw data with criteria and rules that it needs to match so the algorithm will find a structure and information that it can follow. In other words, it will find patterns and similarities by making mistakes. This model is most suitable for complex datasets and for identifying patterns in the dataset that are not obvious. The algorithm will itself not understand the patterns, but it will cluster and group the data into different sections after analyzing the dataset. Unsupervised learning can be divided into three different categories. (geeksforgeeks,"Unsupervised Learning", 2023)

- Clustering, most common
- Association rules
- Dimensional reduction

3.8 Clustering

Clustering is when the dataset is divided into groups that have similar data and features, so it will analyze the dataset and find some characteristic patterns and similarities to put them into different groups. This is done without any prior knowledge of the labels. The main reason to use clustering is to partition all the data into subgroups. (Priy, " Clustering in Machine Learning",2024)

3.9 Semi-supervised learning

Semi-supervised learning is a combination of supervised and unsupervised learning. This model is based on both labeled and unlabelled data. The distribution is a small amount of label data and a large amount of unlabelled data for it to train with. The purpose of this model is also to use historical data to predict an accurate prediction of the result. The semi-supervised learning will treat the label and the unlabelled data differently, for the label dataset the algorithm will update the model weights and for the unlabelled data, it will minimize the difference in prediction by training the dataset. (Bewtra, "The Ultimate Guide to Semi-Supervised Learning",2022)

3.10 Reinforced learning

Reinforcement Learning is a self-learning algorithm with a feedback system, it consists of an agent that is self-trained in an environment. It's for the agent to learn the optimal behavior in the environment by doing and getting a reward when doing a good job, and punished when doing wrong. The goal is always for the agent to take the best action to achieve the maximum punishment. Over time and after a learning period it will be able to make the optimal decision and take the optimal path to achieve the maximum reward. Reinforced learning uses an algorithm that applies a trial-and-error method and receives feedback to learn. (Bajaj, "Reinforcement learning"2023)

3.11 Online learning

As mentioned before traditional machine learning takes a historical dataset and learns from the dataset that is given by us the humans. Therefore, it has a limited dataset that we choose to give it access to. But Online learning is different because it gets and updates the dataset while the process is working, it has a continuous workflow that feeds the machine learning with new data and will update the algorithm and retrain it when the new data has been collected. The whole process of new data coming into retraining and making a new prediction takes milliseconds.

Traditional machine learning gets a batch to learn and the large data set that is given from historical data will get trained to predict future predictions of the actions. But as time goes on the prediction can get worse since there are new types of actions that have never been seen before. Therefore, there is online learning that will adapt to the new types of actions that appear in the environment. It updates the understanding of the patterns and retrains the algorithm. Although is efficient and has a real-time update process the cons of this algorithm are that it changes the model every second and that can affect the accuracy and precision rate due to that it is not possible to control what is being inputted. When it updates the dataset every second it's a high risk for overfitting and the model will start to "forget" the old information. (Pagels," What is Online Machine Learning?" 2018), (Awan," What is Online Machine Learning?", 2023)

3.12 Production line

An assembly line is a process of production that breaks down manufacturing into smaller steps to complete the product. This is usually applied to the mass production of automobiles to increase the working time and make it more efficient. An assembly line is a series of individual workers assembling a product, where everyone is performing a specific task in a particular sequence. Applying this to automotive manufacturing which has been done will increase the number of produced vehicles. (Banton, " Assembly Line: Defining the Mass Production Process",2022), (Britannica," assembly line" 2024)

An automotive assembly line is divided into many different sections, it starts with the bare chassis, and components are attached accordingly during a moving line also called a conveyor. Parts are being sub-assembled on the lines beside the main line to streamline the process of assembly of the parts. The operators are those that perform a specific task along the converter with a part and tools that are necessary. An intricate system of scheduling and control ensures that the parts match the model of the vehicle. Simultaneously as the operators work there are machines to facilitate the assembly, which are often supervised by humans. Many processes and assemblies are, however, still being performed manually, because it's too complex or too hard for machines to reach. (Banton, " Assembly Line: Defining the Mass Production Process",2022), (Britannica," assembly line" 2024)

3.13 Tightening process

In a car production line, there are several different screws and processes to apply depending on the car part and model. This leads to the need for different tools for the different screws and parts to regulate the force and size. One of the tools that are being used during the tightening of a CC1 screw is an Atlas Copco screwdriver, figure 3.4: (Atlas Copco, n.d.) "Angle Hold and Drive Nutrunner LTV . The screwdriver has different settings that are being programmed before usage. A car has multiple parts that need to be assembled and one of the ways is to fasten them together with one or multiple screws. The process is to place the screw in the pre-drilled hole and then fasten the screw with the screwdriver that will spin until it reaches the right amount of torque and angle.

To get a better perspective of the screw and tool that can be used, there is a screwdriver in figure 3.4 and a type of screw that can be used in the production line, figure 3.5.

Figure 3.4: (Atlas Copco, n.d.) “Angle Hold and Drive Nutrunner LTV”

Figure 3.5: Screw used in production

3.14 Previous research

From previous research “Process curve analysis with machine learning on the example of screw fastening and press-in processes”, the research was focusing on the analytic area of the process of a screw fastening and press-in process with the help of machine learning. By using ML to make a curve monitoring the aim was to detect and classify errors in screw fastening. The research acknowledges that the preload force is the most crucial indicator to ensure that the screw does not get loose or too tight then it can break. (Meiners et al., 2023)

By working from a graph-based analysis they had different checkpoints where they controlled so the screw was in place. The first check was the threading phase to ensure that the press-in was on the right path. Then there was a transition check to see if the curve runs in a defined manner, with consideration of the top and bottom limits. Then the last check was to monitor the block force of a press in process if it would have an envelope curve. The graph analysis was based on the force(N) on the y-axis and displacement (mm) on the x-axis. (Meiners et al., 2023)

The aim was to investigate the curves using different ML which was done by recreating different errors to show how they can detect errors in the fastening. They compared different MLs on their made-up data set that was created by a handheld screwdriver from DEPRAG SCHULZ. According to (Meiners et al., 2023) the fastening was done by pressing the lever at the shaft of the screwdriver until the screw was fastened. They divided their incorrect fastening into different subgroups and had one OK fastening group. In total, they had 228 fastenings and 47 of them were OK and the rest were incorrect fastenings. (Meiners et al., 2023)

According to (Meiners et al., 2023) they mainly tested AdaBoost, gradient boosting, SVM, random forest, and MLP. The values were directly fed into the ML and were trained 30 times and had an 80/20 percentage split for test and train. The result that they came to was that by using a curve monitor with ML on the screw fastening process based on torque curves they could detect different types of error on the incorrect fastening, the most suitable algorithm was 1D-CNN. They also mentioned that the naïve Bayes classifier achieved a precision of 100 % and had no classifying. (Meiners et al., 2023)

4 Workflow process

This chapter dives deeper into how the work process looked, from the first ideas to the end concept, and how it changed along the way. The difficulties and the reasons behind why the result did become the way it is.

4.1 Project outline

To be able to study and research the positive and negative impact, a project such as "remove manual borging" is a perfect project to be able to gain a better understanding of how AI and ML can be used in production. Volvo Cars is in the end looking for an in-house product that they own and can implement in their production line. The product should also work as a two-verification tool to keep high safety and assure quality, but also to be able to provide the necessary information/data. The data that the external workers provide when performing the borging on screws.

The program should work with PowerBI for visualization for the engineers at the offices to read and to be able to detect faulty screws that cannot be seen by the human eye. But also, compatible with working beside the machinery and processing a lot of data quickly since the workers in the factory work on a time frame.

4.1.1 Current situation

The machines that are used in the production line work with software where certain PKI values are programmed into the tools in the station and assure that the machine can give the right output for every screw and follows the constraints, if not the system will mark that as red which means error. The current workflow is shown in Figure 4.6, starting from the production site.

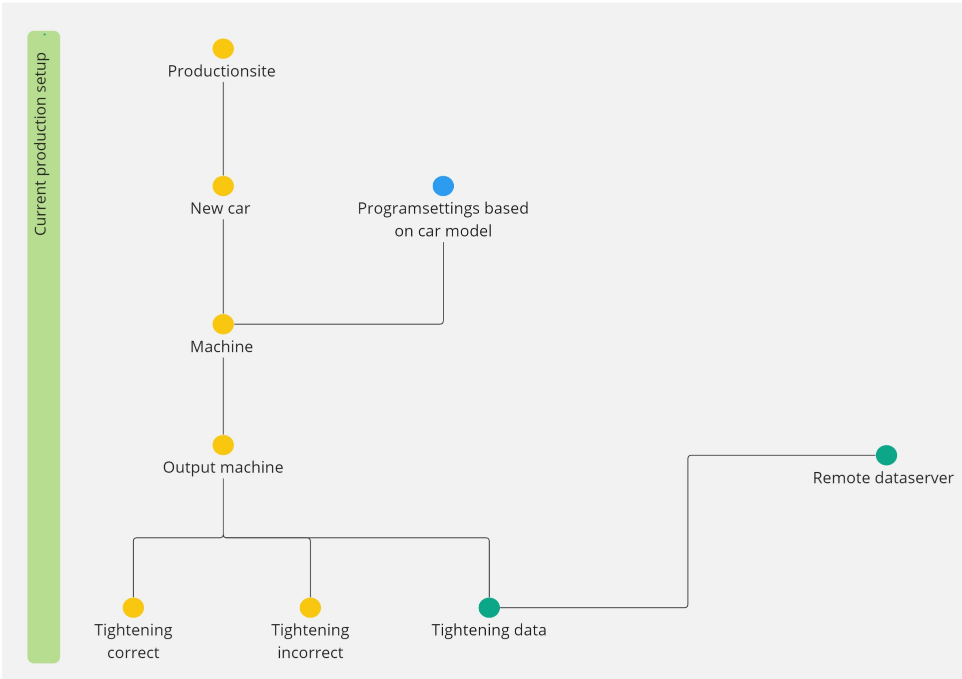

Figure 4.6: Current workflow for tightening in production

The machinery's software can't be tampered with or adjusted with other systems, which makes it necessary to create software that can work beside it. At the same time, it is important that it can be corporate with the machines. Each machine does register some values during the full tightening process, torque [Nm], time in seconds, and angle [deg]. After the screw process is done and if it is correctly done it gets reported to a database and marked as "OK". On the other hand, if the tightening has some problem during the process, it will get reported as "incorrect". That database is accessible by Volvo Cars personnel and can be viewed for a set period and then the visual element gets removed. In the production line, the operators will see if the tightening has been done correctly with a display shown in Figure 4.7. The data will be stored in software and it is possible to see the graphs as shown in Figure 4.8.

That data can be downloaded locally and placed in an Excel document for further review, along with data such as "Program name", "car model", "which station", and a lot more.

4.2 Project strategy

The project started with gaining a full understanding of how the full process works, what is possible, and what each value means. Through discussions with people linked to the different

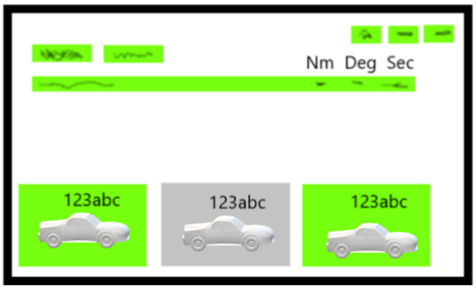

Figure 4.7: Display in production showing if a tightening is correct or not

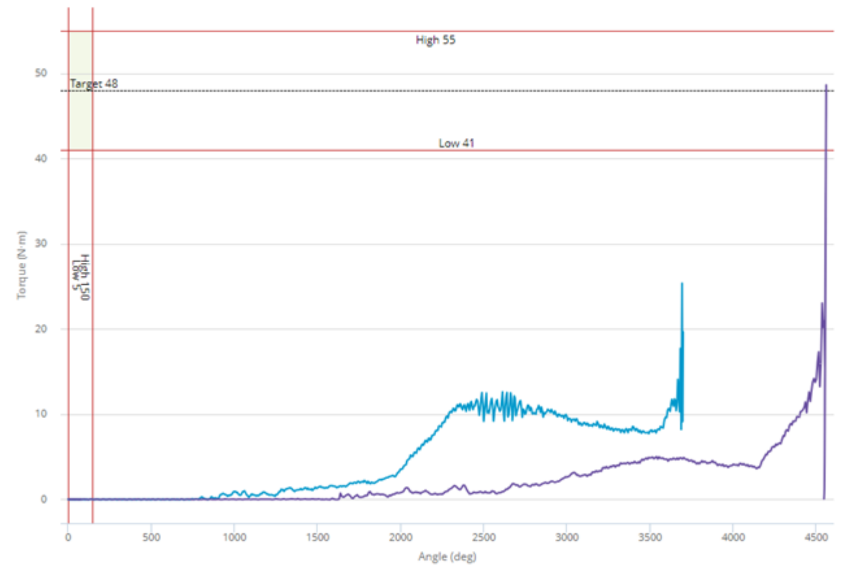

Figure 4.8: “OK” and “incorrect” tightening graph

parts, an understanding started to form. Our main questions at the initial state were:

- What is borging?
- When is borging done?
- Why is borging performed?
- How is the borging process performed?
- Screw types, which is most common?
- What is a PKI?
- Needed features?

These questions were important since it’s necessary to know the problem and the current state to be able to fix it. Also, it’s very important to know the purpose and the value of the process and the needs and requests.

Then the manual analysis of the data began to find focus points, weak points, irregularities, and interesting patterns. Defining the needs of the final product was required to narrow down the research and applicability. A conference call with a company that already has a similar product to gain an understanding of how they tackle a similar task.

4.2.1 Brainstorming

When most of the requirements and features were defined a brainstorming process was started. The focus was on how to contain the graphic movement of the torque by the angel. Ideas were created and discussed, and many of them continued in the process, but some were also scrapped.

Idea 1

An initial thought, before a full understanding was gained was the idea to use contour lines to create an above and below the limit which the torque by time can move between to have a correctly tightened screw. But that thought was marked as not useful. That is due to the machine's software as mentioned earlier can't handle additional parts. That's how a standalone software was determined needed.

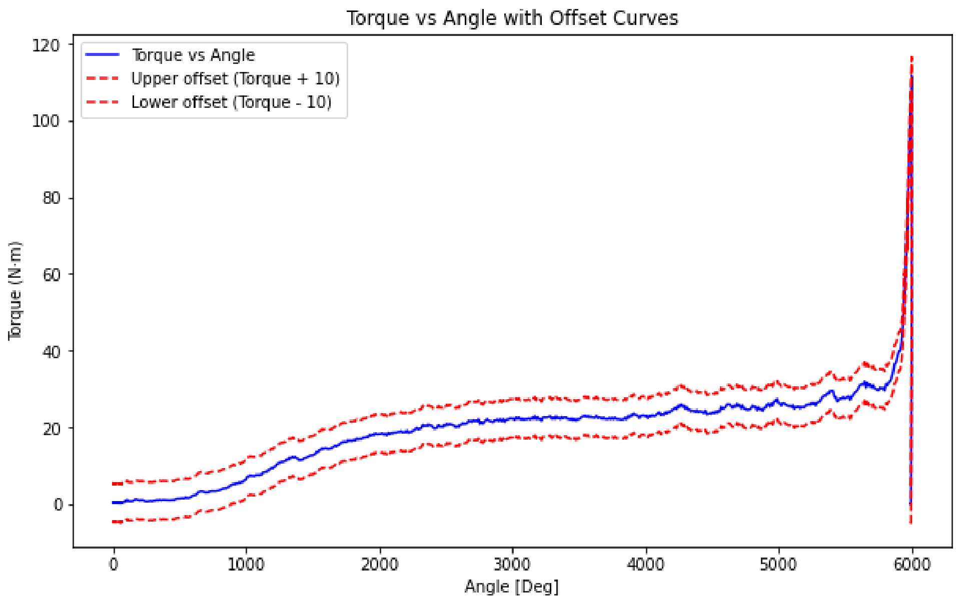

Figure 4.9: Graph with upper and lower limits

Idea 2

By focusing on online learning which is a subset of AI which is found in chapter 3.11. Initially, interest was sparked in the possibility of utilizing a feedback loop. Also to be able to train the dataset directly in the assembly line. However, since it is a complex environment to work within and there is not a lot of accessible research in that field. It was put into consideration but not in the focus of this project.

Idea 3

Machine learning was the focus. Through research and some testing, the use of machine learning did match better with the product needs and in a simpler manner. Since ML can use offline learning, it can use a bigger database which is necessary for the ML to gain an understanding of the patterns. For it to later determine if a screw was correctly tightened or not. Somewhat simplicity of ML against AI, the ease of the possibility of integration, and less need for personal resources further down the road at least initially made the use of ML for this project a smarter choice.

4.2.2 Electing algorithm

The choice of working with ML meant a decision of which ML algorithm needed to be done. There are a lot of different machine learning algorithms, all equipped with different sets of tools to be able to tackle different problems, issues, or assignments. Finding which to use based on your needs is crucial. That process does need a lot of comprehensive research into a bunch of different algorithms that work like each other but handle data differently. So, the algorithm should be able to provide a result based on your provided data but also handle the data correctly, so that the picking process is delicate.

4.2.3 Data analyzing and data cleaning

Each tightening process had a couple of hundred rows of data, which made the need to understand the purpose of each row of data important. Could we reduce the data? Compromise it? Where a couple of questions did appear when the algorithm-picking process did start.

4.2.4 Research viable algorithms

Based on what outcome that was searched and the needs of the algorithm, a comprehensive research phase was started to find and narrow it down to five possible usable algorithms for this exact assignment. Through researching and analyzing the available database and the purpose of every algorithm and matching the knowledge together with the availability of resources. The first decision was to narrow it down to use supervised ML and a classifier since it's a binary problem.

Neural networks represent a powerful approach in machine learning, particularly adept at handling complex patterns and large datasets. Unlike traditional algorithms like logistic regression or decision trees, neural networks mimic the structure and function of the human brain, composed of interconnected nodes (neurons) organized in layers. Each neuron receives input, processes it through an activation function, and passes it on to the next layer. Through iterative training on labeled data, neural networks adjust the weights of connections between neurons to minimize errors and improve accuracy in predicting outcomes.

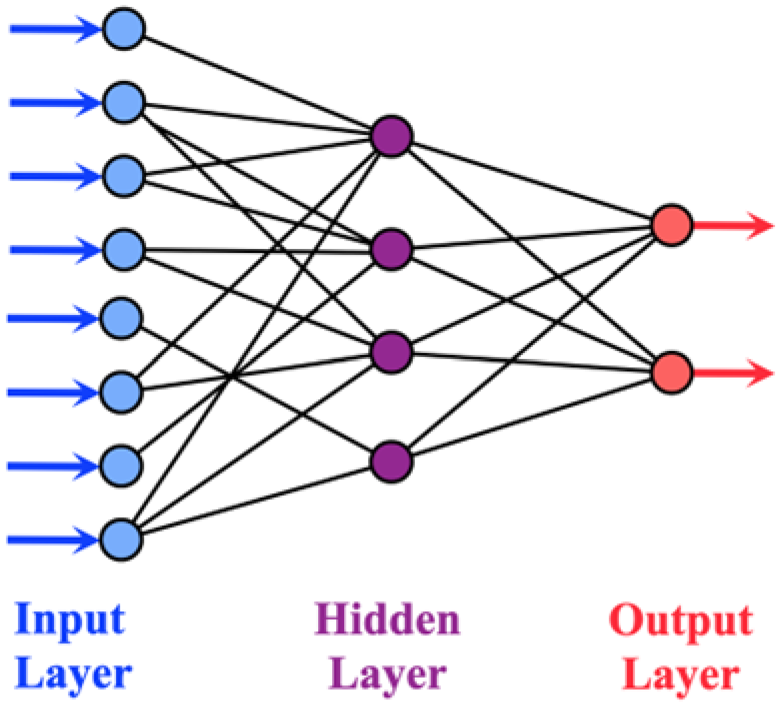

Figure 4.10: Visualisation of how neural network works, (Shallow Neural Networks, 2024)

Figure 4.10 offers a visual depiction of how a neural network operates, illustrating the flow of information through interconnected layers. This visualization showcases the network's ability to capture intricate relationships within data, making it a compelling choice for various machine learning tasks. As we delve deeper into the research phase, we'll consider the applicability and performance of neural networks alongside other algorithms like logistic regression, random forest, decision trees, support vector machines, and k-nearest neighbors to identify the most suitable approach for our specific assignment.

Through a narrow search, the following five algorithms were chosen to proceed further to the "final" chosen algorithm to start creating the code that can solve the assignment.

- Logistic Regression
- Random Forest classifier
- Decision Tree
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)

4.2.5 First boxplot

A boxplot, also known as a box-and-whisker plot, is a graphical representation of the distribution of a dataset. It displays a summary of key statistical measures, including minimum, first quartile (Q1), median (second quartile, Q2), third quartile (Q3), and maximum. When analyzing and researching five different algorithms with the same data, the following result is given. In Figure 4.11 a boxplot shows clearly which algorithms are best performing.

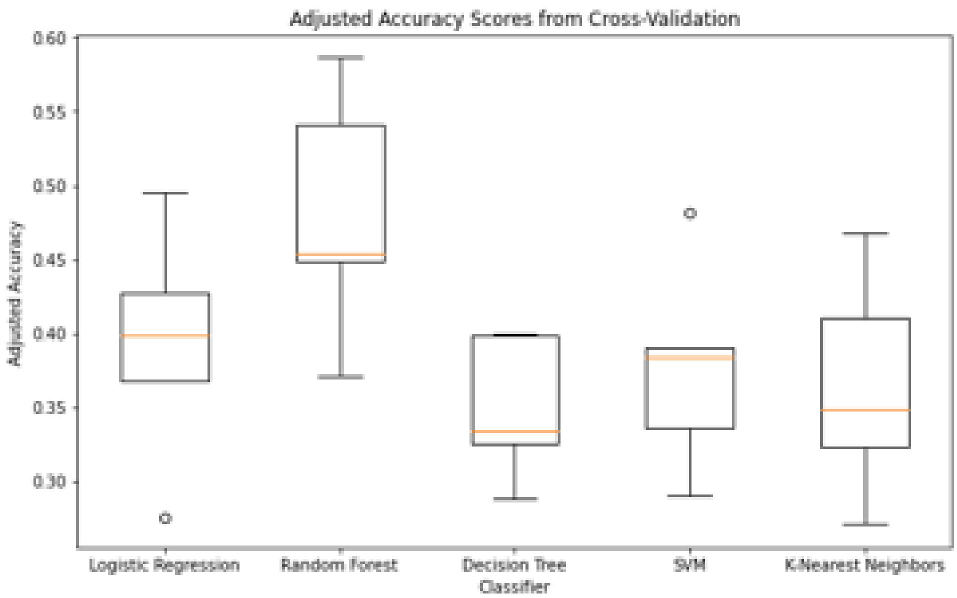

Figure 4.11: Boxplot one to see best performing algorithm

4.2.6 Adding searched for features

Some of the features that are searched for by Volvo Cars and the current manual process are needed for the software as well. Rules are defined and needs are being fulfilled. The most sought-after features were developed. By applying the new program the process will look as in Figure 4.12.

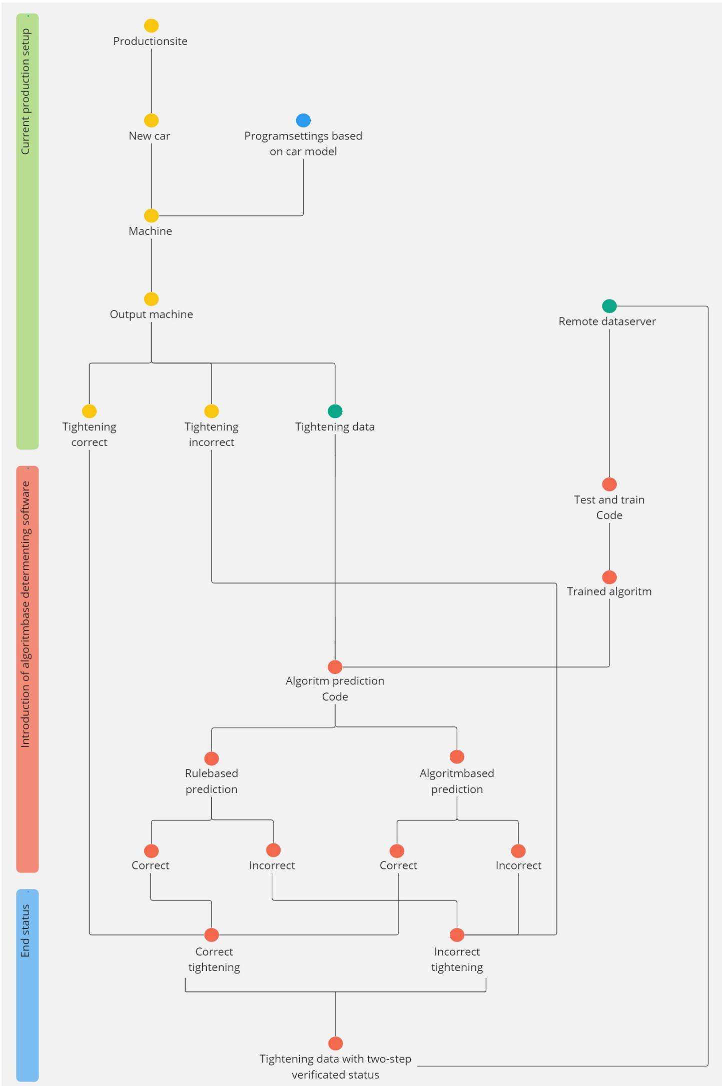

Figure 4.12: Workflow with software added

Provide PKI

PKI are values that the machines currently check the tightening against. Values that are needed by the screws to be measured up to. Target values define the different steps in the tightening process. Minimum and maximum are values that should be the value it needs to surpass or be below when the target value is reached.

New recommended PKI values:						
	Torque [Nm]			Angle [Deg]		
Conditions	Target value	Minimum value	Maximum value	Target	Minimum value	Maximum value
Condition 1	*According to standard*	-	-	-	-	-
Condition 2	*Set standard value*	-	-	-	*Code calculated value*	-
Condition 3	*Code calculated value*	*Code calculated value*	*Code calculated value*		*Code calculated value*	*Code calculated value*

Table 4.2: New recommended PKI values

Visualization

The visualization is important due to being able to ensure the software and the machines do not let incorrect tightening pass through which does in the end create faulty cars, which is unsafe for the customers.

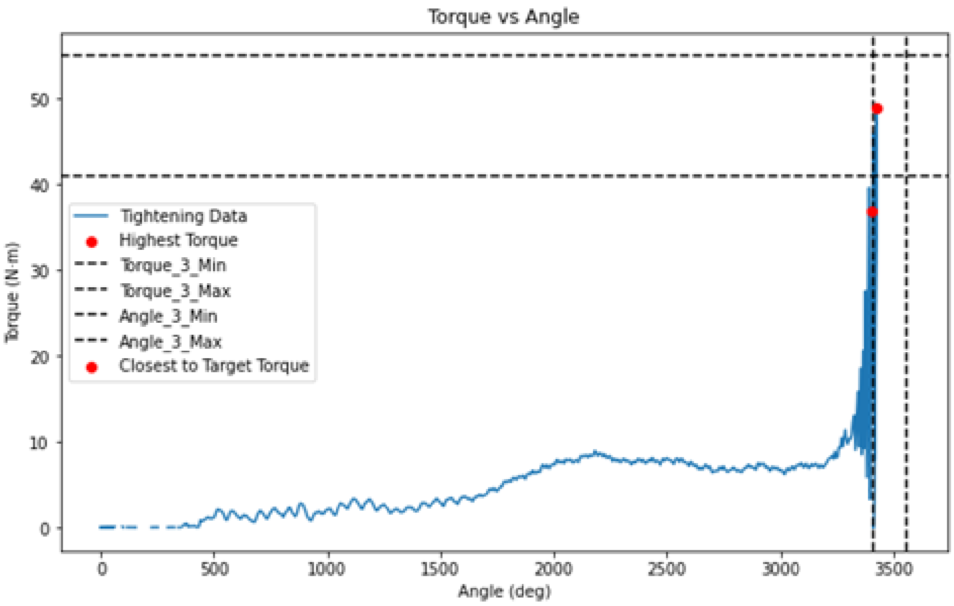

Figure 4.13: "OK" tightening graph

Result prediction

Getting an accurate result is one of the most important things since it determines whether the cars will be let out to the customer or not. At an initial stage, the machines will work with the software as a two-step verification process, which means the car needs to pass both the be accepted and move forward in the production process.

The output of the result for a given screw with its identification number	
Overall prediction for Identification number "*Identification number*":	"*OK*" or "*NOK*"

Table 4.3: The output of the result

4.2.7 Scope out

After conducting a deeper analysis and gaining a comprehensive understanding of the data and the entire code, a "zoomed-out" view was taken to explore potential improvements and adjustments aimed at achieving better results. Also by looking through the whole code in detail and then reorganizing the data. The same comparison of the five different algorithms was executed and in Figure 4.14 there is a boxplot that provides better results.

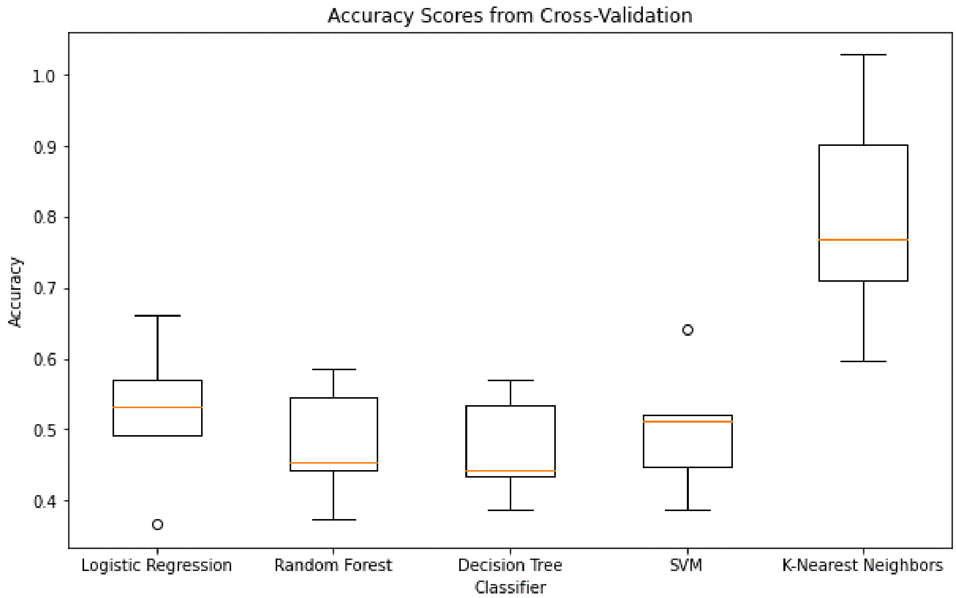

Figure 4.14: Boxplot two to see new best performing algorithm

Implementing the new best-performing algorithm K-Nearest Neighbors (KNN) created a big growth in accuracy.

5 Result

This chapter will present the outcome of the work and the final concept will be presented, as well as how to use and apply the program to the production line.

As mentioned before the borging is being done manually by many different operators to ensure the safety of the cars. The solution to replace the manual work with AI and machine learning has been accomplished by analysis and programming.

The final concept is a program that can make predictions on the status of a new tightening that has been done by a screwdriver. The program is based on a selected dataset, that has been handpicked by us to achieve a good variation of incorrect and correct status on historical data. The algorithm that has been used in the program is KNN and is based on a training and testing split. The algorithm will analyze the historical dataset in different ways. Then based on that it will later, when new tightening data is coming in make a prediction and put out a status of either 1 or 0 where 1 is OK and 0 is incorrect To ensure that the safety remains the same or higher, there is a two-step verification. Before it even arrives at the predictions state, the new tightening goes through the rule state, in this step, multiple rules have been added.

The rules are based on PKI and it's 3 different checkpoints that the tightening must achieve to continue in the process. The 3 different checkpoints are based on the torque and angle values and must coincide at a certain point, the checkpoints are the green boxes in figure 5.15. The program will recognize what type of screw the new tightening is based on the program name. The program name will be in the historical dataset and the new tightening. The first checkpoint is located at the start where the torque must reach a value and at the same time, the angle can't be under or above another value. The values are based on the PKI which is based on the screw and their properties. If the new tightening is stopped in one of the checkpoints it will be marked as NOK in the status. But if it goes through and matches all the checkpoints it will continue to the prediction state as described above. Points system have been added to verify the safety of every tightening, It's a presentation of the point system in tables 5.4,5.5 and 5.6.

Rule-based prediction	Points
OK	1P
NOK	0P

Table 5.4: Rule-based prediction

ML-Prediction	Points
>50%	1P
>75%	1P

Table 5.5: ML-prediction

Overall prediction	Status
1P AND 0P	NOK
2P	UNCERTAIN
3P AND 4P	OK

Table 5.6: Overall prediction

The historical dataset only needed to be processed once since it is being saved to ensure that the accuracy is and will remain the same for future predictions, this process takes around 100 seconds with the current dataset. Therefore, the prediction process for one tightening will never take more than 1-3 seconds to ensure that is possible to use in the production. Every station in the production line has a time limit to follow to be able to keep a good flow in the production line otherwise it will stop every few seconds and they will not be able to produce all the cars in time.

The final concept is a working program that has a two-step verification in the prediction state to ensure safety for all the cars that leave the factory. First, it goes through the rule-based prediction and if the status is OK then it goes further down the workflow to KNN algorithm-based prediction, which has historically achieved an accuracy of 87%. This is possible due to that the dataset has been cleaned and analyzed before letting it be worked on in the program. Volvo Cars also expressed interest in fine-tuning their PKI values, so incorporating this into the program was essential. An analysis of the historical dataset was conducted to potentially generate improved values that could be applied effectively. Figure 5.16 is a representation of the workflow.

Figure 5.15: Graph with three checkpoints

All printed after prediction stage	
Searched for information	**Printed information**
Model loaded correctly	
Initial acquired accuracy from test and train process:	"Acquired percentage" %
Overall prediction for Identification number "*Identification number*":	*OK*" or "*NOK*"
PKI table	
Visualization of tightened screws curve	
Execution time:	"*Runtime prediction stage*" seconds

Table 5.7: The output of the prediction code

Model loaded assures that the screw tightening data is correctly loaded into the prediction code. The initial acquired accuracy is from the code that tests and trains the algorithm gets printed with the percentage value along with two decimals.

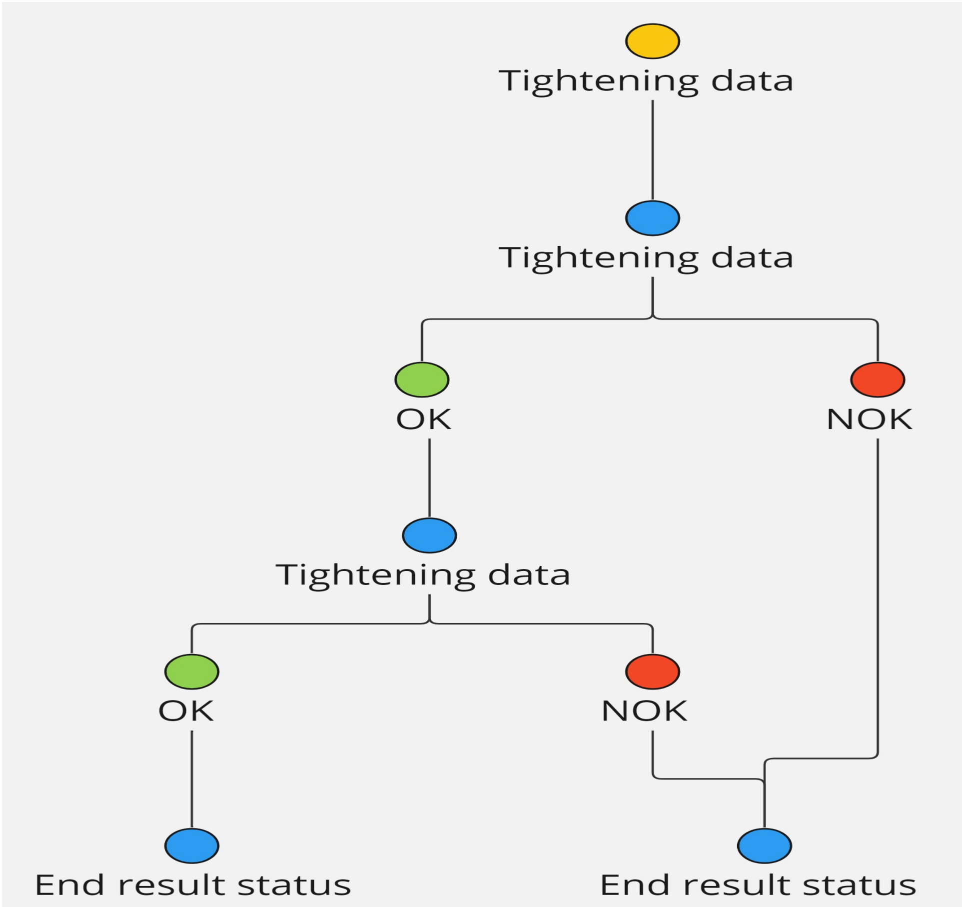

Figure 5.16: Workflow of the prediction

5.1 Test and verification of the program

To ensure that the end concept will work, multiple tests and verification were made both during and after the process. After letting the algorithm work in a train and test environment with the cleaned dataset until it got an accuracy that was satisfied by us, a new tightening with the dataset of torque parameter, angle parameter, and the program was inserted into an Excel file and then inputted into the program. The new tightening was then going through both verifications. First the rule-based and then the prediction-based. But depending on the status of the predictions it either continues or stops in the first step. Figure 5.16 shows how the two-step verification works.

5.1.1 How to utilize the result?

There is a code that can store and predict safely with a two-step verification to not let any faulty screw through. The next step for Volvo Cars or other production lines is to connect their dataset directly to the code. To be able to do this it's necessary to develop software, to connect to their available computers on the production line. The software needs to be able to store and take data directly from the tightening process and process it in the code. This is a step for the software department to do. When it's possible to store large amounts of data and be able to connect it directly to the screwdriver it's necessary to test in a test environment so that everything runs smoothly.

The next step is when everything works and we know that the program can work fast and accurately, it needs to work in the background in the production line and see the environment, also to see if it will disturb or negatively affect the process, the deployment will be done from the automation department. If any modifications or adjustments need to be done this will be done by the software or data scientists department due to the knowledge of programming.

Along the way, many tests and validations need to be done and divided into smaller steps to safely be able to pursue the next step. There is a solid start to work on and it is important to know that it takes time to apply this to the production line but it's possible.

Although it should work on most of the stations it's necessary to work and focus on 1 specific station at the start. Just to not overcomplicate it and not to overwork the program. To be able to see if the program can safely work instead of the human this can be tested in the test environment first and when it's ready for launch it will need to work along the original process. After time and analyzing the result, it will be possible to replace the manual work and can be tested in the production line.

Even if the main goal was to replace the manual work for a specific process the research has resulted in a code, that can be applied to daily work routine to ensure that the tightening process is correctly executed even when it's not a borging process. So it can run in the background when all the tightening is being executed and alert the tightening that the current machine does not alert about.

6 Discussion

This chapter will consist of all the reasons for all decisions that have been made during the processes and the thoughts behind every decision.

Why this choice?

Every decision that has been made is based on research, testing, and verification to ensure that it's the most optimal choice. The biggest decision that has been made is the selection of the algorithm. After researching all available machine learning algorithms, a handful of options were considered. Following further investigation, the decision was made to use the random forest classifier for this project.

This decision was made without any bigger consideration of the dataset and more based on the math and how the algorithm works, by using a decision tree. After that a specific station and screw were selected to have a dataset to work on. The choice was important to be able to narrow down the problem, the chosen station was 1.6-798 it's a belt screw in the cars. After experimenting and testing the algorithm with the dataset, we instantly knew that there was something that must be done and it was to analyze the dataset due to the big amount. Today the machine only controls 3 different points which means it only checks for 3 different values in the torque and the angle. 1 tightening has multiple values and rows in an excel file, about 700-800 rows of data from both the torque and the angle. If the ML algorithm checks every value, it would get confused due to every tightening looks different due to the human factor that can affect the tightening process.

The decision of cleaning the dataset was made, only collecting the data when the screw reaches the first value of the PKI, then it will start to analyze and work with it in the test and train environment. 1 tightening has an ID number that contains multiple rows, therefore, it needed to be grouped so the algorithm knows that the 1 tightening ID has multiple rows. The reason for this was so the data from the same tightening stay together in the test and train environment, before getting shuffled.

Now when the algorithm is chosen and the data is cleaned some adjustments need to be made, after testing different scenarios, it lead to the fact that the random forest classifier was no longer the optimal algorithm. Therefore, further investigation was made on some other algorithm and KNN became the most optimal algorithm for this specific case. During the whole process, Numerous tests and verifications were conducted to identify ways to improve the system and ensure its applicability across various stations and screws in the production line. The goal was to develop an adaptable solution, within the manufacturing process rather than being limited to a specific screw or station.

Other outcomes?

The initial thought was to program a code that could create an over and underline and adjust itself by reading in all the data, so ML learns the pattern and creates the optimal pattern to follow. The issue with this thought was that it would create a problem when the ML is working and take more time to read and analyze the dataset and then create the graphs for it to follow. In other words, it's too time-consuming, and in the production line every second counts. We also discovered that it will be almost impossible to connect to the machinery that is available today. This initial idea would have worked like a stock chart and predicted the outcome of the data that came in. Although this idea would have given the operators a good perspective of the graph it would have taken a lot of time for it to work in the environment and the operators would not need to see the graph, only the status. Therefore, this idea was not an alternative to continue within the later stages due to the inconvenience. Figure 6.17 shows the idea.

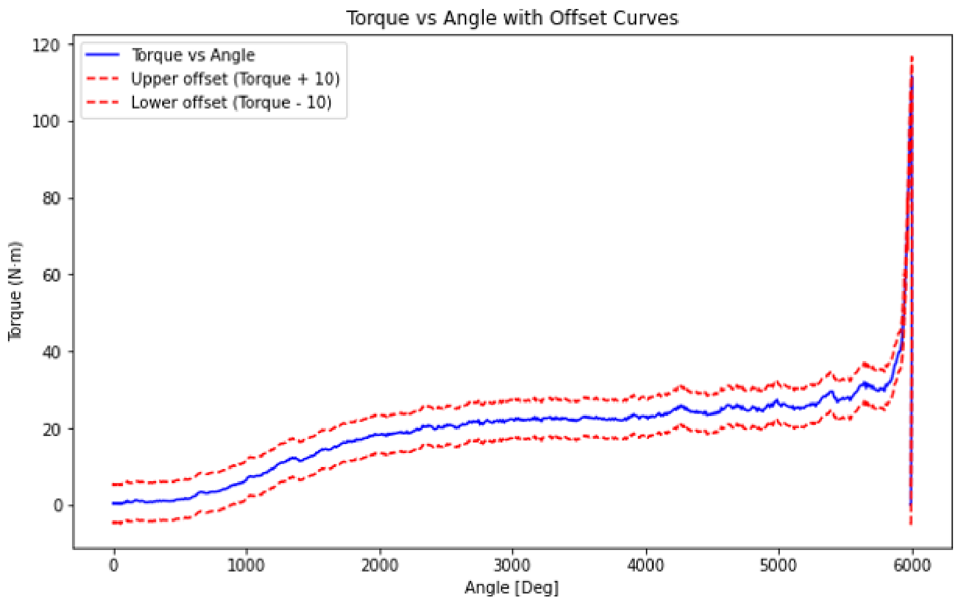

Figure 6.17: Upper and lower offset on a tightening process

Another thought that we worked on for a long time was to add a feedback loop to the code so if the prediction from the ML code is wrong, the operators would have the opportunity to correct it. Also for the ML code to remember and store that mistake so when a similar pattern comes in it will have learned. This idea needs to be investigated more and was something that the time frame restricted us from from. We think that this idea is something necessary and good for development so the ML algorithm can constantly grow and learn from the mistakes that it has made. Because every tightening is different due to the human factor and can not be removed. We investigated the feedback loop and realized that it's possible to save the learned information in a file but not how to transfer the information to the algorithm without going into an online environment so we decided to be our limit. Figure 6.18 gives a visualization of what we did and want to continue to develop.

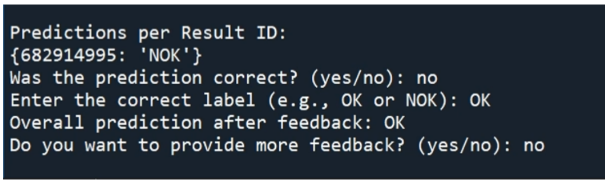

Figure 6.18: Python code of the feedback loop

The third small idea that did change in the later stages was to change from the random forest classifier. We always knew that it was a binary problem therefore we chose to focus on the available classifier. The main focus was on the random forest classifier because it could handle a big dataset. The random forest classifier checks almost every point and makes multiple decisions for it to analyze train and test in the environment. But big is not always the best, although it covers a lot of data handle it, it did have some flaws and the time for it to analyze a big portion of data was too long because it needed to analyze every single point of the dataset in multiple different ways to learn and test. The result of this algorithm did not satisfy us, and we decided to compare other alternatives that could give us an accurate result fast enough to be applied in the factory. Therefore, the decision to go with another algorithm was the right choice.

If we did not keep an open mind, we may have been stuck in the early stages or not have gotten a result that we would be satisfied with. The important thing is that we discovered new ways to go around the problem and look at it from different perspectives. This is possible since more knowledge was gained and if we had chosen an alternative solution, it might have worked but we always tested and did research before continuing with a new idea.

What could have been done differently?

Every decision that has been taken has been evaluated and compared with alternative solutions. Still, it's important to be able to look back and see if anything could be done differently. We are satisfied with the outcome, especially due to the time frame and available resources. In the process, during this time there is nothing that we would change. However, looking back at the process if we were more open to trying more algorithms in the beginning there would have been more time in the end to test and optimize the algorithm. But at the same time, we would not know if we would gotten the same result that we have accomplished today. Therefore, we see this as a future implementation and everything that we also wanted to add and test but could not be due to the short time and resources will be considered in the future.

Customization

It's possible to apply this program to similar stations in the Volvo Cars factory without any modifications to the code. The only thing that needs to be adjusted or loaded is the new dataset from other tightening. If a new station or factory in Volvo cars wants to apply this the code doesn't need any adjustments. Only the dataset needs to be updated and let the program analyzes and work with it. This has been tested on 1 other station and gave a positive result, with high accuracy and reliable result. This program is in the early stages and needs more work and can therefore not with safety say that it applies to all stations and gives high accuracy. More tests and validations need to be made. But in the current station 1,6-798, every operator can use the program and get the same outcome, since the dataset is handpicked and carefully selected. This also means that every station can not only just apply the random amount of dataset and get the same accuracy, at the moment it needs to be handpicked and tested before being applied.

Future implantation's

Before diving into future implementations it is important to keep in mind that the test and training code and the prediction codes are developed in a scaled-down environment. The code has been tested on existing tightening data that are available but now needs to be connected to the tool. Making up tightening situations to test the code is the next step and finding a way to automate the workflow between the downloadable data from the screwdriver to the code directly.

Currently, it is not known how exactly the production sites of the future will look, will there be operators like today or are they replaced with machines? Either way, implementing and using software like this could help with forming the future of production, by increasing safety and efficiency along with lowering the production cost, with customization according to needs and budget. Utilizing AI and ML to ensure vital things through analyzing lots of data points in different environments quickly and efficiently is the correct way forward. However, in the future implementing AI and ML in some way into any production me be very likely. As in the Volvo Cars case, it is easier to find a faulty tightening when studying 450 data points than looking for two to four searched-for values.

Research of online learning and adding a feedback loop has been considered, to be able to add a feedback loop would be very useful. Then the borging process when all the external operators check multiple screws on every car that comes, could be replaced. The new borging process will in the beginning consist of 1 operator in the station. All the tightening will be assembled as usual and, in the background, there will be a program which is the rule and ML-based prediction program. When all the screws are tightened and its time for borging only 1 operator needs to be available to only check the incorrect tightening that the program has warned about, and this will be sustainable and safe for the cars that roll out of the factory since it consists of a two-step-verification. Therefore, instead of checking all the screws on the multiple cars, it's now only necessary to control the screws that the system alerts about with the help of the ID number. This will lead to a lower labor cost when it's possible to reduce multiple operators to only 1 supervisor to stay ready by the computer.

Here is when a feedback loop needs to be added although the program can safely detect the incorrect it can also sometimes, predict an incorrect when it is OK. The other way around is almost impossible since there is a two-step verification. Therefore a feedback loop can be necessary for the ML to learn from the mistake and store this information. The feedback loop would be placed in the computer next to the production line and when a control of an incorrect tightening is done, a response needs to be applied whether it was a correct prediction or not. If it was a correct prediction nothing going to be stored, but if it was an incorrect prediction, a response needs to be sent to the program so it labels the tightening with the data and places it in the right category. This is achievable with the help of online learning, but there are some negative parts of online learning, and it's also why we chose to continue to go offline.

Online learning will learn from itself and create a database on its own, it's possible to choose how often the database should update. In this particular case all the new tightening will be stored in a way and later keep adding on to the main data frame that will go through the test and train stage. This means that the accuracy can be changed during this update and in the long term the old data frame will be forgotten about due to the large input of data. It will update every second or more often and the split of the classification will not be equal so that it will be more OK than incorrect due to that more OK tightening in the production line, it will affect the test and training split so the model will have less incorrect to learn and work with. It's better to have traditional learning and handpick the database and retrain it manually when it's time and when a long-term problem accrues in the production line. Although there are negative parts online learning can be applied on a smaller scale and we can limit the intake of the new data and restrict it to have a certain split and add the feedback loop so it is possible for it to learn by doing.

This project will provide a more efficient and fewer errors when the human factor is deflected. The human operators have many factors that can affect the working quality, which can be sleep, energy, and the surrounding environment that can be described. When the borging process is being executed, visual control of the tightening position after assembly. Check if the screw head is flush with the base of the component. Perform a torque check with a click wrench and then mark it with a pen. To repeat this on multiple screws and multiple cars will affect the focus and the accuracy of the human. Therefore, by replacing it with AI that does not get affected by the environment it can save time and mistakes, the most beneficial saving is the cost of labor when it is possible to replace x amount of time and x amount of workers with 1 superior that has an overall view of the situation.

Another future implementation to add is PKI recommendations that provide the team with better PKI values to set for the machine and for it to follow to get a smaller checkpoint area to let less faulty tightening through. This is something that is half implemented in the code, the code can provide it but only missing the math behind it. So now it only gives us hypothetical numbers.

Something to also keep in mind is that we had multiple conversations with our supervisor at Volvo Cars to ensure that we stayed on the right track. Over time when more knowledge and a better understanding were gained, some aspects changed, we went from "Remove manual borging" and changed our goal and target to automating or semi-automating parts of the production line with the help of the new technology, ML, and AI.

7 Conclusion

In conclusion, there is a fully functional code that can predict in 1-3 seconds based on the chosen dataset. To assure that every car that leaves the factory has a fastened screw was crucial therefore, there is a two-step verification that will not let any faulty screws through the system. The optimal machine learning algorithm for this specific case is KNN due to its ability to carry a big database and make fast and accurate predictions. In this binary case, it can adjust to every station that involves a screw and a screwdriver in the Volvo cars production line.

The outcome of the research is that the use of AI and more specific machine learning is possible in a production environment. This will contribute to a lower cost for the company since they can minimize the external operators that they have to take in to do the borging process. When this process is automated the time and cost can be reduced. By automating the process, the human workload will decrease, and this will lead to the human factor being something to disregard. Often the human factor is a crucial reason for the fault that occurs in the factory and that will affect the cars that roll out of the factory. Quality assurance will increase when the human factor is replaced by machine learning and AI.

To answer the main question of this thesis "Is it possible to remove the manual borging and replace it with machine learning and AI?". Yes, the process can be replaced with an automated system that in the start will need to be supervised to ensure that everything works well. The code will need further development but is today good enough for the early stage, it can analyze data and adjust to different screws and make a safe prediction. The borging process needs to remain but can have another workflow. A workflow that is beneficial for both the operators and the company. The same purpose will remain, to ensure that every car that leaves the factory is safe to drive. Although this project was for a specific quality assurance step in the factory it can be applied for daily use, it can be applied as a second layer of security for the existing screw that the current machinery can miss.

The research was done on Volvo cars production line and mainly focused on the pre-trim department in manufacturing. Although this code is customized for this case, the research has provided us with a broader knowledge of AI and ML. Therefore this can easily be adjusted and applied to similar cases both in the Volvo Cars production but also for other production lines and other tightening and fastening processes. The potential for increased efficiency, reduced costs, and improved quality assurance makes AI and ML indispensable tools in the evolution of production processes.

References

1. *Our heritage — Volvo Cars.* Volvo Cars.
   https://www.volvocars.com/intl/v/our-heritage

2. VolvoCars. (n.d.-a). *The three-point safety belt - over 1 million lives saved.* https://www.volvogroup.com/en/about-us/heritage/three-point-safety-belt.html

3. VolvoCars. (n.d.-b). *This is Volvo Cars.* Volvo Cars Ingress Swede Automobile.
   https://volvo.ingress-swede.com.my/this-is-volvo-cars/

4. Wikipedia Contributors. (2024, October 24). *Volvo Cars.* Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Volvo_Cars

5. Atlas Copco. (n.d.). *Angle Hold and Drive Nutrunner LTV - Pneumatic Assembly Tools — Smart Portal.* https://www.atlascopco.com/enuk/itba/products/assembly-solutions/pneumatic-assembly-tools/pneumatic-nutrunners/ltv/ltv48-had-angle-nutrunner

6. Bajaj, P. (2023, April 18). Reinforcement learning - GeeksforGeeks. GeeksforGeeks.
   https://www.geeksforgeeks.org/what-is-reinforcement-learning/

7. Bansal, S. (2024, March 13). Supervised and Unsupervised learning - GeeksforGeeks.
   https://www.geeksforgeeks.org/supervised-unsupervised-learning/

8. Banton, C. (2022, April 8). Assembly Line. Investopedia.
   https://www.investopedia.com/terms/a/assembly-line.asp-0

9. Bewtra, A. (2022, July 1). The Ultimate Guide to Semi-Supervised Learning.
   https://www.v7labs.com/blog/semi-supervised-learning-guide

10. Britannica. (2024, March 18). assembly line Definition — Britannica Money.
    https://www.britannica.com/money/assembly-line

11. Google. (n.d.). What is Supervised Learning? Google Cloud.
    https://cloud.google.com/discover/what-is-supervised-learning

12. Gupta, M. (2024, March 20). ML — Linear Regression - GeeksforGeeks. GeeksforGeeks. https://www.geeksforgeeks.org/ml-linear-regression/

13. Keita, Z. (2022, September). Classification in Machine Learning: A Guide for Beginners. Datacamp. https://www.datacamp.com/blog/classification-machine-learning

14. Meiners, M., Mayr, A., & Franke, J. (2023). Process curve analysis with machine learning on the example of screw fastening and press-in processes. https://www.researchgate.net/publication/349198101_Process_curve_analysis_with_machine_learnng_on_the_example_of_screw_fastening_and_press-in_processes

15. ML — Types of Learning - Part 2. (2023, December 4). GeeksforGeeks. https://www.geeksforgeeks.org/ml-types-learning-part-2/
16. Pagels, M. (2018, April 20). What is Online Machine Learning? The Hands-on Advisors. https://medium.com/value-stream-design/online-machine-learning-515556ff72c5
17. Priy, S. (2024, March 20). Clustering in Machine Learning - GeeksforGeeks. GeeksforGeeks. https://www.geeksforgeeks.org/clustering-in-machine-learning/
18. Shallow Neural Networks. (2024). Questions and Answers in MRI. https://mriquestions.com/shallow-networks.html
19. What is the Difference Between Lazy and Eager Learning? (2024, February 13). GeeksforGeeks. https://www.geeksforgeeks.org/what-is-the-difference-between-lazy-and-eager-learning/
20. Abid Ali Awan, What is Lazy Learning? (2024, February 13). GeeksforGeeks. https://www.datacamp.com/blog/what-is-lazy-learning
21. Abid Ali Awan, What is Eager Learning? (2024, February 13). GeeksforGeeks. https://www.datacamp.com/blog/what-is-eager-learning
22. Abid Ali Awan, What is Online Machine Learning? (2024, February 13). GeeksforGeeks. https://www.datacamp.com/blog/what-is-online-machine-learning