



Vietnamese-German University

COPYRIGHT WARNING

This paper is protected by copyright. You are advised to print or download **ONE COPY** of this paper for your own private reference, study and research purposes. You are prohibited having acts infringing upon copyright as stipulated in Laws and Regulations of Intellectual Property, including, but not limited to, appropriating, impersonating, publishing, distributing, modifying, altering, mutilating, distorting, reproducing, duplicating, displaying, communicating, disseminating, making derivative work, commercializing and converting to other forms the paper and/or any part of the paper. The acts could be done in actual life and/or via communication networks and by digital means without permission of copyright holders.

The users shall acknowledge and strictly respect to the copyright. The recitation must be reasonable and properly. If the users do not agree to all of these terms, do not use this paper. The users shall be responsible for legal issues if they make any copyright infringements. Failure to comply with this warning may expose you to:

- Disciplinary action by the Vietnamese-German University.
- Legal action for copyright infringement.
- Heavy legal penalties and consequences shall be applied by the competent authorities.

The Vietnamese-German University and the authors reserve all their intellectual property rights.



Big data analytics and its application in
production quality
and worker management


BACHELOR THESIS Vietnamese-German University
Binh Duong, 2022

Submitted by: Tran Thi Khanh Linh

VGU Student ID: 14957

Supervisor: Dr. Dinh Hai Dung

Big data analytics and its application in production quality and worker management

Submitted to the Department of Mechanical Engineering of the

RUHR-UNIVERSITÄT BOCHUM and VIETNAMESE-GERMAN UNIVERSITY

in partial fulfillment

Of the requirement for the degree of

BACHELOR IN MECHANICAL ENGINEERING

September 2022



Tran Thi Khanh Linh

Mechanical Engineering

Binh Duong, 09th Sept 2022

Non-disclosure note

Internal material from two companies is contained in this bachelor's thesis, which are anonymous in the literature as "company" or "organization." Without the express consent of the aforementioned company and author, third parties may not access this work, except for the thesis supervisors and authorized examining individuals. Unless the companies and author expressly consent differently, the material provided in this document would not be shared publicly until the final bachelor's viva voce examination.



Vietnamese-German University

Abstract

In today's data-driven economy, transforming data into actionable analytics is the most effective way to increase efficiency, quality, and productivity. Because of rising demand, manufacturing processes are becoming increasingly complex.

The current customer-centric and data-driven business world, along with evolving social and technological trends, continues to drive product quality and human resources to new heights. Big data is providing significant value to manufacturers. Automation, quality, production reliability, and information analysis are all used in the latest multidisciplinary fields of big data technologies in the production industry. While in the human resource sector, big data greatly supports planning and analytics.

Thus, the thesis gives a general review of big data analysis along with its application in a manufacturing company environment, specifically regarding product quality and people.



Vietnamese-German University

Acknowledgment

My gratitude goes to my supervisors, Dr. Dinh Hai Dung in the BIS program and Dr. Nguyen Hong Vi in the GPEM program. Both of them have offered me unwavering support, constructive criticism, and invaluable insights into a fast-paced research field. I appreciate both of you bringing me on and providing me with an incredible chance to study with you during this whole process. You have always been an influence on me, and I look forward to working with each of you again in the future.

Many of my colleagues and classmates at the Vietnamese-German University as well as in business have also backed me up. I appreciate all of the debates and general encouragement I've received concerning my research.

My heartfelt sincere thanks to my friends and family, especially Le Minh Quang. Without your consistent support, this journey would have been drastically different: thank you for trusting in me and encouraging me in all of the strides I've chosen to take.



Vietnamese-German University

Table of contents

Non-disclosure note	2
Abstract	3
Acknowledgment	4
Table of contents	5
List of Abbreviations	8
1. Introduction	9
1.1. General background	9
1.2. Problem definition	10
1.3. Research questions	11
1.4. Purpose	12
1.5. Delimitation	13
1.6. Thesis structure	14
2. Literature review	16
2.1. Big Data Overview	16
2.2. Big data analytics	18
2.2.1. Big data analytics	18
2.2.2. Data analytics in manufacturing	19
2.3. Quality management in production	21
2.3.1. Product quality	21
2.3.2. System of quality management	21
2.3.3. Lean initiatives in manufacturing	22
2.3.4. Lean Six Sigma	22
2.3.5. Related researches	24
2.4. Worker planning	24
2.4.1. Purpose	25
2.4.2. Strategic worker planning with support from data analytics	26
2.4.3. Components	27
2.4.4. Related projects	28
2.5. Data-driven decision-making approach	30
2.5.1. Developing critical questions	31
2.5.2. Data gathering and analysis	31
2.5.3. Informing decision-makers of the results	31
2.5.4. Refinement of procedures, organizations, or systems	31
3. Application in a production quality management research	32
3.1. Research Strategy	32
3.2. Case study	32
3.2.1. Background	32

3.2.2. Approach	34
3.2.3. Proposed impact	36
3.3. Method for data collection	37
3.3.1. Dataset scope	37
3.3.2. Data sources	37
3.3.3. Data collection methods	37
3.4. Method for data processing	38
3.4.1. Machine-generated data processing	38
3.4.2. Human-generated data processing	38
3.4.3. Statistics processing	38
3.5. Method for data visualization	40
3.5.1. Power BI	40
3.5.2. Microsoft Excel	41
3.5.3. Microsoft PowerPoint	41
4. Application in a worker planning project	43
4.1. Project Strategy	43
4.1.1. Waterfall project management	43
4.1.2. Agile project management	44
4.1.3. Project timeline	44
4.2. Conceptual design	45
4.2.1. Technical Design	45
4.2.2. User experience	46
4.2.3. Proposed Impact	47
4.3. Method for data collection	49
4.4. Method for data processing	51
4.4.1. Data pre-processing	51
4.4.2. Data processing tools	51
4.4.3. Back-end processing	52
4.4.4. Front-end processing	55
4.5. Method for data visualization	55
4.5.1. Worker selection by skill	56
4.5.2. Training assignment	56
4.5.3. Edit information	57
4.5.4. Upload/Download resources	57
5. Result	58
5.1. Production Quality Management	58
5.2. Worker Planning	60
6. Discussion	62
6.1. Big data analytics	62
6.2. Production quality management project	62

6.3. Worker planning project	62
7. Conclusion	64
8. Bibliography	65
9. Appendix	67



List of Abbreviations

JSON	JavaScript Object Notation
CSV	Comma-Separated values
HTML	HyperText Markup Language
SPC	Statistical Process Control
FMEA	Failure Modes and Effects Analysis
DoE	Design of Experiment
HR	Human Resource
DDDM	Data-Driven Decision-Making
UAA	User Account and Authentication
DB	Database
SSO	Single Sign-On
JPA	Java Persistence Application programming interface



Vietnamese-German University

1. Introduction

1.1. General background

Data is around us. It has assisted businesses in making better decisions years after years. Adopting big data solutions has greatly benefited businesses from all industries. In fields including healthcare, energy, banking, telecommunications, marketing, and sports, this is true. The industrial sector has also benefited from the digital era, significantly boosting any nation's GDP. Manufacturing is essential for national economic systems because it stimulates innovation and the creation of jobs. It is widely known that this industry is the most challenging and complicated in terms of the range of products.

Manufacturing data is frequently used for operational purposes by manufacturing organizations. Big data applications in manufacturing include:

- Preventive and predictive maintenance

The availability of machinery in the manufacturing process is critical to operational efficiency. Manufacturers can use analytics software to determine the machine's condition.

- Automation of Production Management

This is likely the most complicated use of big data in production.

The case of sensor technology in wind turbines is an excellent example of production management automation. Sensors collect data on energy generation and wind direction, allowing the wind turbine's blade pitch to be adjusted to maximize efficiency.

- People Service

A satisfying customer experience has become crucial to every organization. Big data analytics tools gather information from a variety of sources, including clients, staff, and pertinent stakeholders, which could be used for extensive data processing.

1.2. Problem definition

Data does not appear on its own. They are always acquired and processed by being "algorithm ready", "scrubbed", and "cleaned" for use in complex algorithms. As a result, human decisions about selection, judgment, interpretation, action, and filters are always influencing them.

In terms of big data, the manufacturing industry faces its fair share of difficulties. Among the difficulties were the following:

- Complicatedness.
Adoption of new technologies is always difficult due to the complex nature of manufacturing.
- Product design and advancement
Manufacturers need to invent frequently to thrive and stay relevant.
- Use with Caution.
Big data frequently causes managers to abdicate judgment and place an undue emphasis on the data. Data must be utilized carefully, particularly in regard to sensitivity and security.

In the practical industrial context for the scope of this thesis, for the big manufacturers, a big amount of structured, semi-structured and unstructured numeric values, documents, images, etc. coming from all sources requires significant effort to be put into the collecting, processing, and visualization phases before being able for the stakeholders and managers to make an impactful decision regarding people, process, and tools.

1.3. Research questions

The research was conducted from the standpoint of data and information that are accessible, utilized, and significant for the industry and people, as well as the flow of information in the operational processes, in accordance with the nature and the scope of this research. Therefore, the questions for this bachelor thesis were;

Question 1: What are big data and the corresponding analytics?

Question 2: How are big data and its analytics currently utilized in enterprise production, specifically in quality management and labor resource coordination?

The goal of the author's bachelor thesis is to categorize current research, find relevant topics and trends in the literature, and identify intra-industry resources that are directly related to big data analytics in production industries. The primary research question may appear plain because the purpose of this is to create an overview that illustrates the major factors linked with the topic . As a consequence, the following is the thesis's key research discussion:

"What is big data analytics and its role in manufacturing and human resources?"



Vietnamese-German University

1.4. Purpose

The goal of this thesis is to systematically study how big data and information are employed in manufacturing systems as well as human resource information tools for operation. This research is conducted in an academic and industrial context.

The purpose of this research's conclusions is to help standardize the process for the next quality management system and workforce optimization. Determine the importance of information management in an organizational sense, moreover.

This thesis' major goal is to examine big data, big data analytics, and their roles in the system of organization management as a whole. Two particular goals that correspond with the structure of this paper might be created from this general purpose:

The first one pertains to the theoretical aspect and describes the knowledge that concerns big data analytics. Hence, the author will begin by properly describing the concepts of big data, quality management, and workforce planning. In a later section, the author will discuss how it is connected to corporate decision-making, concentrating on how it is applied throughout the many stages of a manufacturing organization. The author will next examine how it affects the organization and reveal what options are available to prevent and address the associated issues.



The second relates to the practical aspect and uses two examples to demonstrate the value and function of big data in business. The cases were the author's actual work and documentation intra-industry. The applications will be implemented using both internal and open-source software, and they will be modeled on the various daily operations that the business conducts. The management of the collected data will be the base for the analysis.

1.5. Delimitation

All of the study's discussions center primarily on data and information management. In this thesis, "data" is defined as figures, facts, or symbols, "information" is defined as data that has been analyzed and processed in a way that makes sense . Data is necessary for the production of knowledge and information . Although most studies frequently use data and information interchangeably, as they will in this thesis, However, it is crucial to understand the main difference between the two terms. Therefore, the focus of this study is on the management of data and information in the assembly data systems for the creation of manual instructions inside two manufacturing businesses (referred to as "Company" or "Organization" anonymously). More so, the empirical data for this study was gathered within these plants, the confidential data were omitted and replicable data was replicated. Since big data analytics is the main focus of attention as well as its two implementations in studies and projects in the manufacturing industry, other functions and processes within the supply chain will not be taken into account and will therefore be delimited.



Vietnamese-German University

1.6. Thesis structure

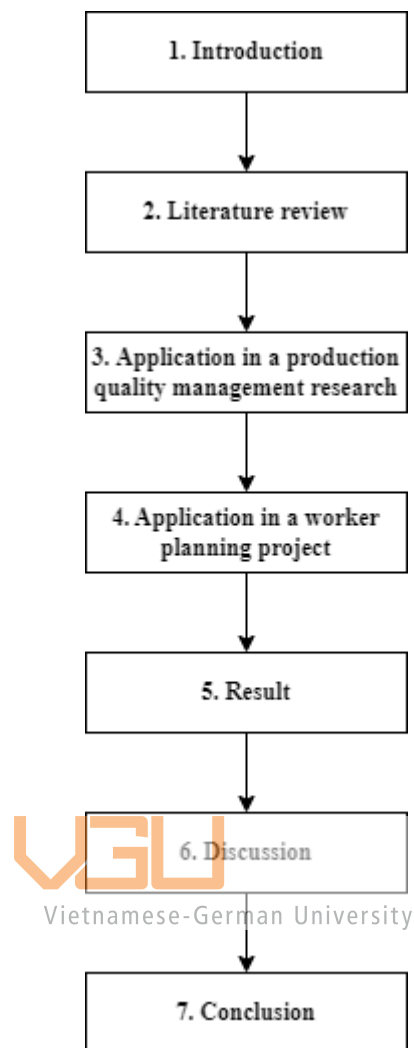


Figure 1.1. Disposition of thesis

The issue statement and thesis questions are established in Chapter One, **Introduction**, which also exposes the reader to the topic of the thesis.

The reader is given access to the academic research done on the subject in Chapter Two's **literature review**. Rather than conducting a comprehensive search, this chapter identifies the most important and relevant studies on the issue.

Chapter Three, **Application in a Production Quality Management Research**: describes the exact processes utilized to carry out this application of big data analytics in quality management and describes the research methodology and design that have been employed to conduct it.

Chapter Four, **Application in a Worker Planning Project**: describes the exact processes utilized to carry out this application of big data analytics in human

management and describes the project methodology and design that have been employed to conduct it.

Chapter Five, **Result**: In this chapter, the study's findings are reported. The presentation of the conclusions from the data gathering process in Chapters Three and Four is a significant aspect of the authoring of this chapter.

Chapter Six, **Discussion**: discusses results and draws inferences about the study's findings. The discussion of the findings in connection to the problem statement and research questions defined in Chapter One is an important aspect of authoring Chapter Six.

Chapter Seven, **Conclusion**: draw the final conclusions to the thesis.



2. Literature review

2.1. Big Data Overview

The massive, difficult-to-manage volumes of both structured and unstructured material that continuously overwhelm enterprises are referred to as "big data". But it's not only the kind or volume of data that matters; it's also how organizations use the crucial data. Big data might well be investigated to provide insights that strengthen conclusions and give assurance for crucial business decisions.

	Traditional data	Big data
Database architecture	Centralized	Distributed
Data types	Structured	Structured , semi-structured, unstructured
Data volume range	Gigabytes to terabytes	Terabytes, petabytes and beyond
Data schema	Fixed or static	Dynamic

Table 2.1 Comparison between traditional and big data

Several scholars have addressed the three key Vs of big data features and their importance in the realm of big data analytics. Initially, differentiation was based on three key characteristics: volume, velocity, and variety.

Volume is concerned with the enormous size of data that is constantly rising for each sector; it is enormous and allows improved forecasting for the future.

The **velocity** of data refers to how quickly it can be analyzed to make decisions (access); data is increasingly arriving in large waves, and it is critical to obtain useful analytic information in real time.

Variety describes the incredibly diverse resources of data (structured, unstructured, and semi-structured) at the schema level, such as word, sensor, sound, video, chart, and more. As a consequence, a variety of data are gathered, allowing for a more thorough analysis of large data.

Because of the rapid technological advances, the characteristics of Big Data analytics have grown over the years. As a result, researchers have addressed this issue by reconsidering the three Big data criteria and adding new ones.

The term "**veracity**" refers to the numerous values, which differ significantly in terms of data coverage, accuracy, and timeliness. Because extremely important business choices were made based on those data sets, it is critical to guarantee that the data set under consideration is adequately correct.

Value is defined as an individual's or an organization's ability to convert large data into real values, which includes the ability to acquire and then harness data to achieve specific goals.

The presenting of data of nearly any sort in a graphical manner that makes it easy to understand and interpret data is referred to as **visualization**. Choosing the incorrect visual aid or merely defaulting to the most frequent kind of data visualization could lead to viewer confusion or incorrect data interpretation. Understanding what sorts of graphs to use and why is one of the challenges that slow down reporting and analysis.

Vulnerability refers to the security measures that must be put in place to ensure that the data acquired is treated in compliance with the law and the customer's requests.

The data kinds of big data were classified as follows based on the sources.

- Data that is structured
- Data that is unstructured
- Data that is semi-structured

2.2. Big data analytics

2.2.1. Big data analytics

Large amounts of data may not always imply high quality. And when numbers increase, it becomes increasingly crucial to concentrate on quality in order to extract some valuable insights from the current data. The usability of the information, which is often evaluated by its value, continues to be a key factor in determining whether a company must make an investment in big data or not.

Advanced analytical methods are used on large data sets in big data analytics. It hence focuses on big data and analytics, and how the pair have come together to deliver one of the greatest important advancements in knowledge management today.

Deep data analytics restructures the company's behavior in a variety of ways. Data analytics uses approaches such as information technology, data, quantitative techniques, and numerous other methods to achieve success.

A thorough understanding of traditional analytics is required in order to evaluate the factors that are leading many firms to Big Data analytics. Structured data sets are a component of traditional analytical techniques that are routinely assessed for particular objectives. Large data sets are frequently taken into consideration while developing the most of data tools. In fact, it is a basic principle that the bigger the dataset, the more precise the statistics as well as other diagnostic conclusions will be. Rather than applying quantitative and mining methods, the data engineers tend to construct and/or manually write elaborated SQL that handle enormous amounts of data in quest of the perfect stakeholder group, or exorbitant operating expenditure. The most modern management information systems and analytical features both employ big data. Popular data models for organizing and processing business applications include the relational model. There are tables in this system with distinct names where relevant data is kept in columns and rows. They can also quickly parse tables and run large queries. We have reached a new level of performance that is particularly compelling for applications requiring large data thanks to recent generations of industry tools and platforms.

There were 3 major kinds of analytics.

- Descriptive analytics serves as the foundation for more advanced analytical models. This version responds to the question, "What happened in the enterprise?" It pays leaders, managers, and planners to create a guideline for

future activities by analyzing the database to gain knowledge about current or historical business transactions. This approach does a thorough evaluation of data to provide specifics such as performance expenses, the source of incorrect measures, and the frequency of events. Descriptive analytics aids in determining the source of the problem.

- Predictive analytics provides an answer to the question "What is meant by estimating and comprehending possible challenges via the use of statistical approaches and various forecasting methodologies. It predicts the future using techniques such as machine learning, data mining, and data visualization. The majority of contemporary advanced analytics and big data tools and approaches are relatively tolerant of the transactional schema, non-standard statistics, and low-quality data found in raw source data. That's advantageous because breakthrough and predictive analytics rely on numerous details, including dubious data.
- Prescriptive analytics provides a solution to the question "What is meant by complicated data derived through descriptive and predictive analyses?" This variation chooses the best course of action so prescriptive analytics should be used, for instance, when determining the most cost-effective route to transfer items from one organization to another.

2.2.2. Data analytics in manufacturing

Modern production environments are information environments where information can be shared, transmitted, and analyzed to give manufacturing insight. Manufacturing intelligence may enhance, to name a few things, operational effectiveness, innovativeness, and ecological consequences. To store ever-larger data sets (also known as "Big Data") with the present information technologies providing corporate and operational intelligence enable the real-time calculation throughout this Big Data utilizing complicated algorithms, just as in other sectors and areas. A rise in the number of tools that record readings from settings and operations, as well as a boost in legacy automated processes and sensor technologies, as well as current and innovative concepts will be used to transfer this raw data, which will lead to the anticipated accelerating growth in data production. In the intensity with which these gadgets record and perpetuate measurements, Insights may consume reduced micro information from

various technologies and modeling software, enabling manufacturers to better understand their operations and processes and gain insight that can enhance current ones.

Digitisation, engineering, information systems, and business intelligence are just a few of the multidisciplinary fields of study that focus on big data technologies in industrial contexts. Understanding the present status of study in big data solutions for manufacturing is essential at this time, as is identifying areas where further research activities should be concentrated in order to assist the systems and technologies for the upcoming industrial generation. Therefore, the objectives of this paper is to classify current research initiatives, identify major research themes, and pinpoint any gaps in the body of existing literature.



Vietnamese-German University

2.3. Quality management in production

Production is the act of creating items/products from parts or raw materials, or more simply, the process of transforming raw materials into completed commodities.

2.3.1. Product quality

The dimensions of product quality include:

1. Performance: Will the product perform its intended job?
2. Reliability: How often does the product fail?
3. Durability: How long does the product last?
4. Serviceability: How easy is it to repair the product?
5. Aesthetics: What does the product look like?
6. Features: What will the product do beyond the basics?
7. Perceived Quality: What is the reputation of the company selling this product?
8. Conformance to standards: Is the product made exactly as the designer intended?



Vietnamese-German University

2.3.2. System of quality management

The main components of the quality management system are:

1. Quality Planning: This is a strategic activity that is as important to an organization's long-term commercial performance as product development plans, financial plans, marketing plans, and human resource plans. Quality planning entails identifying clients, both external and internal to the organization, and determining their demands.
2. Quality assurance is a group of procedures that ensures that standards of quality of goods and services are upheld and that problems with quality between customer-supplier are avoided and/or resolved effectively. The reporting of the quality assurance system is a critical component. Policy, procedures, instructions and specifications, and records are the four components of quality system documentation.
3. Quality Control and Improvement: Quality control and improvement refer to the activities performed to ensure that products and services fulfill criteria and are

continuously improved. Because unpredictability is frequently a primary source of poor quality, statistical approaches like statistical process control (SPC) and designing of experiments (DOE) are important tools for quality control and improvement.

2.3.3. Lean initiatives in manufacturing

Lean is a system of practices that concentrates on locating and reducing wasteful activity in any organization. Lean facilitates the production and supply of the proper quantity of high-quality goods and services at the correct time, the first time, while eliminating waste and being flexible.

Any expenditure incurred in an operation that does not benefit the client is referred to as waste. The goal of lean manufacturing is to eliminate waste from production operations.

There are 8 types of waste:

1. Transport: Unnecessarily transporting goods and materials
2. Inventory: Raw materials, work in progress, and final items in excess,
3. Motion: People's needless movements
4. Waiting: Material queuing, idle time
5. Overproduction: Making too much or too soon in comparison to client demand
6. Overprocessing: Incorrect method, incorrect machine, more effort or better quality than required Any non-value-added activity that clients do not want, do not seek, and will not pay for.
7. Defects: Scraps, rework, and components or units deemed unsatisfactory by the customer
8. Underutilized talent: Employee's untapped creativity, motivation, and knowledge

2.3.4. Lean Six Sigma

Businesses can optimize their business processes with the help of the Six Sigma approach. Less faults and higher profits, worker morale, and higher-quality products or services are all benefits of this performance improvement and process variance reduction.

The term "Six Sigma quality" describes a process that is tightly managed

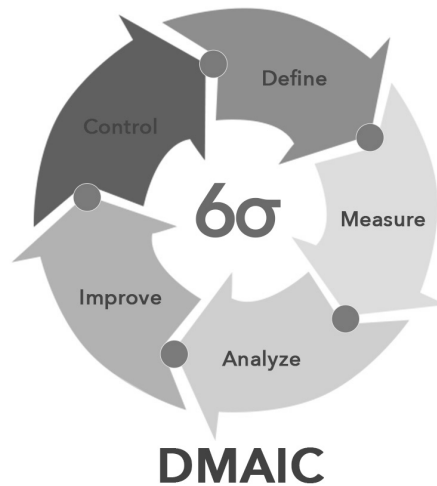


Figure 2.1. 6 sigma dmaic model

A 6σ framework for improving the caliber of outcomes generated by business processes is the DMAIC model. Define, Measure, Analyze, Improve, and Control are abbreviations for DMAIC.

- Define: Identify the problems that need to be fixed and their significance for the company or organization. Organize the group for growth.
- Measure: To describe the fault and acquire baseline information on how well the product or process works. Set improvement objectives and confirm that a suitable measuring system is in place.
- Analyze: Identify the process variables (or inputs) that have the most influence on important process outcomes (outputs).
- Improve: To pinpoint areas that can be made better, and to provide facts to support how these improvements will assist the initiative achieve its goals.
- Control: Implementing the solutions and ensuring their integration into the organizational procedures.

A data-driven, fact-based advancement methodology called Lean Six Sigma places more emphasis on preventing defects than finding them. By reducing variation, loss, and cycle time and promoting job uniformity and flow, it enhances customer

satisfaction and bottom-line performance, giving it a competitive edge. Every employee should take part in it because it applies anytime there is a range of trash.

2.3.5. Related researches

Industry and academic research on big data take different approaches. Industry scientists primarily concentrate on the technical response, solutions, and implementations for Big Data coordination, while academic academics work on the theoretical aspects of Big Data. The majority of academic work is focused on creating new techniques in data analysis, data replication, data distribution, and data handling optimization.

Keqin Wang et al. 2007 suggested one knowledge discovery strategy based on data mining to give high-quality related knowledge. This method can be used to extract information about product quality from a sizable amount of manufacturing data that is relevant to quality.

In the industry, data scientist team proposed the use of the following steps for effective application of big data analytics in defining and testing improvement for quality management:



Vietnamese-German University

1. Prioritize data gathering and analysis by identifying initial patterns (using trend lines, distribution histograms, variance and standard deviation, and clustering).
2. Form a first theory regarding the reasons of yield decline and variability based on the identification of key process performance determinants.
3. Test the initial theory regarding the reasons for variability and a drop in yield, and concentrate your research on the variables that are statistically the most significant.
4. To determine the influence of and ideal ranges for the indicated parameters, model complex processes.

In the research that the author worked at, the mentioned steps were implemented into the case study.

2.4. Worker planning

The company vision has changed to a global vision during the past 20 years. Organizations were being forced to change their business practices as a result of micro

and macroeconomic conditions. Additionally, because of internationalization, competitiveness is getting more and more fierce. The corporation is now under pressure to match its People function's policies with its organizational vision. The HR department's strategies must all be in line with the broader corporate plan. The multiple product and service ranges, service alternatives, and choices that clients and consumers have encourage this paradigm change. Today's undeniable plea is for quality, precision, and cutting-edge technology, and almost every organization adheres to these criteria.

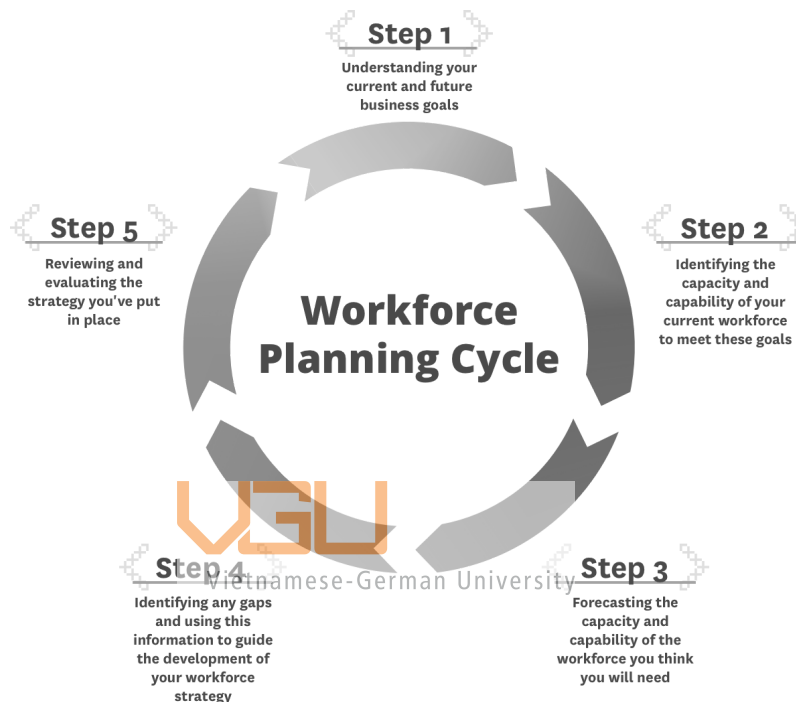


Figure 2.2. Worker planning cycle

Every company's prosperity is seen to depend heavily on its ability to effectively manage its people resources. In reality, in regard to all these other aspects, a company's productivity, staff loyalty, and organizational success depend greatly on its People management Strategic plan.

2.4.1. Purpose

The appropriate people must be in the correct positions at the right times, according to business. As a result, HR professionals need to:

- Employ individuals that fit the company's culture
- Take full use of the individual skill sets that each employee possesses
- Maintain a sufficient number of employees to carry out their company plan

Every firm has a difficulty with managing its human resources. The ultimate objective of a people function's leader has always been to efficiently use employees' untapped abilities and skills. The HR department is critical in measuring its attributes and performance. With the use of modern IT-based solutions, such as big data analytics for the HR department, shortened to HR or People Analytics, which helps in the conversion of huge amounts of data into informative and measurable outcomes, these final responsibilities may now be performed. With the use of HR analytics, multiple different staff management may be done in a more structured manner. Given that human resource departments have long been an essential component of every business, this offers the company an advantage over its rivals.

HR analytics is an empirically supported strategy that combines business intelligence components with tools and techniques ranging from basic matrix analysis to predictive modeling. The purpose of HR analytics is to empower businesses to make wiser strategic decisions regarding their workforce. Analytics that help create a narrative from data are another way to describe HR analytics. The tale is ultimately tied to the overarching business objective through mining the initial raw data for insightful insights. Raw data is processed by HR analytics, which tries to uncover problematic HR indicators and essential areas that need improvement. This helps senior management make crucial strategic decisions.

2.4.2. Strategic worker planning with support from data analytics

Human resources may track projects, absenteeism, monitor and manage calendar assignments, and measure the performance of each employee with the use of HR analytics. The most effective instrument for strategic planning is labor analytics. By identifying cost-savings opportunities, enhancing key talent retention, and boosting workforce productivity and efficiency, HR Analytics is now concentrating on vital staff metrics that relate labor planning to organizational outcomes. This gives HR a seat at the table and the honesty to make company and employees strategic and tactical decisions.

Workforce analytics has become an essential component of HR strategic planning. Today's Human resources leaders base their judgments about the workforce on an understanding of HR analytics. Each choice involving human capital must be carefully considered. As a result, HR analytics assists the HR manager by looking at different

perspectives and making various strategic decisions while considering every aspect of human resources. As a result, Human Resource analytics plays an important role in strategic planning in the following ways:

- Aids in framing what measures should be taken
- Aids in tracking whether or not the implemented solution works
- Aids in tracking the effects of business improvement initiatives.
- Helps to predict certain outcomes that may occur, allowing a strategic plan to be calculated to address the problem.
- Follow the issues in the company, create a strategic response, and assess the effectiveness of the execution.
- Assist HR leaders in identifying and communicating exactly where the company's investment in it's human has high yield
- Helps with long-term labor management by identifying possible talent supply and demand on a local and global scale

2.4.3. Components

To fill in the gap between other operational areas of the organization, various HR metrics must be identified and these metrics measured. The analysis contains the primary component for HR executives to use in developing strategic plans. These elements also serve as parameters for analytics to function correctly:

- Recruitment
The creation of any strategy plan must include this as its most crucial element. Any organization's human resources were a vital and significant component. As a result, businesses make large investments in human resource efforts. In order to optimize the amount of investment in hiring the best person for the job, HR must carefully establish recruitment methods. Analytics therefore helps HR create a precise recruiting strategy that aids in identifying the best applicant for the open post, hence reducing expenses. If the wrong person is selected, the company will incur additional expenses since they will be required to go through the entire selection procedure again, throwing away time and resources.
- Succession planning
Every employee either retires or leaves the organization as time passes. In such cases, an organization must prepare to fill the vacant position as quickly as

possible in order to maintain productivity. It is extremely difficult to predict who will succeed in a key position. The performance of the chosen candidates must be evaluated. Traditionally, It is highly difficult to trace each employee's performance over time. However, analytics help HR monitor the performance of high-performing employees who were able to excel in the aforementioned vacancy. A series of analyses were performed using data analytics, quantitative modeling, and evaluating employee performance over a period of time to help management finalize the promotion and transferring decision

- Performance management

Its main objective is to monitor the advancement of employee productivity as well as the fulfillment of corporate objectives. A well-crafted strategy enhances a manager's capacity to track and comprehend the company's best performers. In order to provide the company a competitive edge, It motivates managers to develop and plan their tactics, establish clear objectives, keep track of progress, and maintain substantial design.

- Training

Identifying training needs assists HR management in increasing employee productivity. The crucial next step, after assessing worker productivity, is to determine the need for training. The firm may improve staff efficiency by developing education programs. Regarding long-term aspects, the employees could have the chance to exhibit themselves better, resulting in better working efficiency and higher income. Due to issues including remote working, resignation, workforce rivalry, and quick technological change, businesses are forced to focus on their employees' learning and development needs.

2.4.4. Related projects

Chandana Siriwardana, 2016 conducted a project to optimize construction labors. Since the construction industry relies heavily on labor, it is essential to organize work crews in order to meet quality, cost, schedule, and productivity goals. The Construction Worker Performance Index (CWPI) was created using thorough worker performance assessment evaluation criteria. To determine the usefulness and applicability of the framework, pilot research was conducted. Following the guidelines of the Delphi survey method, two sets of surveys were used to finalize the CWPI measuring formula.

The development of a worker grading system goes beyond the current traditional labor planning, which simply considers skill levels. An overall framework for worker planning was developed to encompass the whole cycle time of a construction project by integrating various worker planning-related components. Investigations into the interrelationships of many elements led to the identification of traits of both upper and lower-performing labor crews. A correlation analysis was used to look into the connections between the framework's variables. The association between the input factors and the level of cohesiveness among work crews was determined using regression analysis.

This project inspired the worker performance analysis for the project that the author contributed to, which was described in section 4.



2.5. Data-driven decision-making approach

A procedure for making decisions based on data is called data-driven decision-making. With the advancement of data systems and technology, it has become simpler to utilize data to support services

An organization may move to a proof-supported, long run culture with the aid of DDDM. It promotes making decisions using data instead of opinions or instincts, based on experiments and facts. The business switches from being reactive to being proactive. The business uses data to make decisions that enhance initiatives, operations, or plans rather than just reporting for statutory obligations. A company culture built on DDDM rewards choices that are backed up by concrete evidence.

The DDDM process transforms data into knowledge that can be used. Unfiltered, simple facts made up the data. The data can create knowledge that can be utilized to test hypotheses and solve issues throughout time, culminating in the information that informs decision-making, by way of explanations, the supply of relevant background, and the discovery of linkages.

In the DDDM process, data is used to analyze, test, and enhance a system, operation, or plan. There are four iterative phases to these jobs:

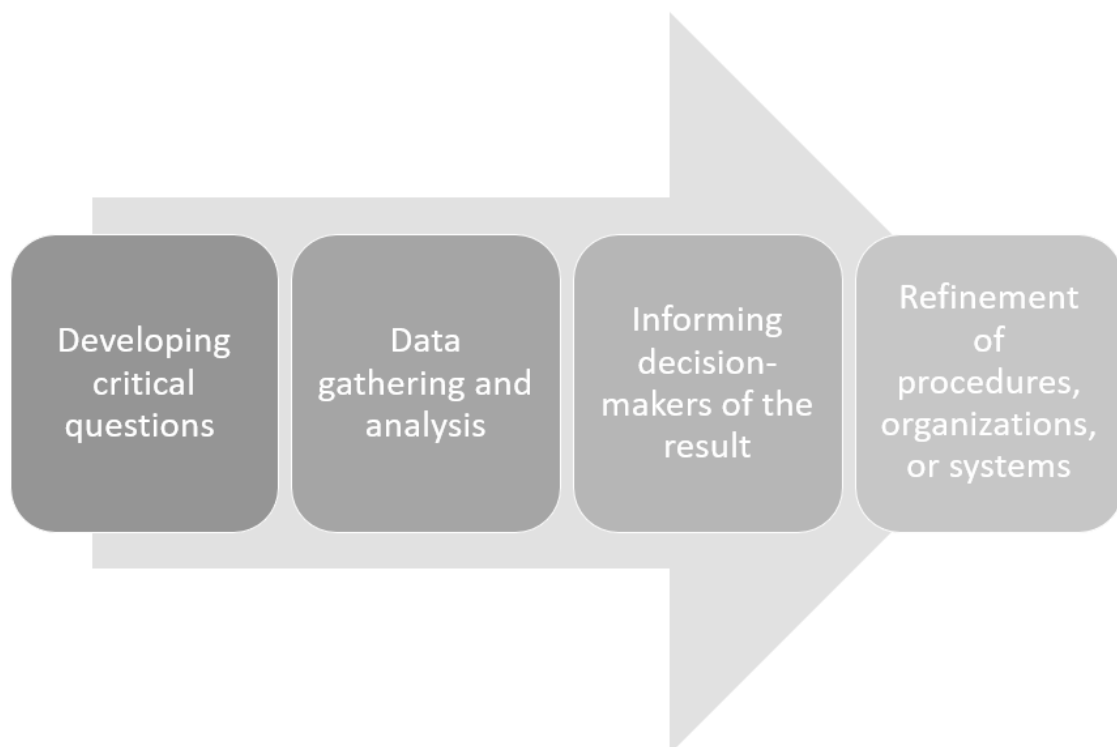


Figure 2.3 Stages of DDDM

2.5.1. Developing critical questions

The approach starts with defining and clarifying the important questions for the organization.

2.5.2. Data gathering and analysis

Using the main questions as a guide, available data is discovered and new data is acquired as needed. It is essential to possess access to quality data.

2.5.3. Informing decision-makers of the results

Major decision at all tiers of the organization or the larger service system were informed of the findings. Dissemination can take many different shapes and take place through a range of communication methods, depending on the contributors' information needs.

2.5.4 Refinement of procedures, organizations, or systems

To discover room for improvement, enhance the effectiveness of programs, businesses, or systems, and evaluate the impact of activities on intended outcomes, they use the data obtained during the preceding stage. As more information is obtained, the process iteratively advances, with fresh data generating fresh understandings and posing fresh inquiries for more information gathering and analysis.

3. Application in a production quality management research

The author conducted this research in company #1

3.1. Research Strategy

A well-designed and acceptable research approach should make it possible to address one or more research questions while also achieving the study objective. Therefore, a research approach is defined by a series of predetermined actions to tackle an issue in order to achieve the study's objectives. Three key factors that were connected to various research tactics can help determine the type of approach which are most effective while doing research.

An in-depth empirical research is called a case study. It is best to inquire as to "how" and "why." In real-world initiatives that were too complex for research or experimental approaches, the case study is generally employed to elucidate causal relationships. Its design involves inquiries or propositions, analytical units, logic linking the data to the inquiries or proposals, and discussion of the findings, just like other research methodologies. A case study may consist of a single case or a group of related instances.

3.2. Case study

The Case study is designed to gather useful information for the decision-makers, using raw and manual data collected from the shop floor.

3.2.1. Background

The case study is implemented in a production line inside a manufacturer's factory with its goal, or optimal quality set as 95% efficiency, However, the overall ratio of passed/qualified product to the total manufactured is mostly in the 90-95% range.

When there are automated failure triggers from the machine through sensor technology, the 2 following scenarios could occur:

1. Random (discrete) trigger: this product failed due to different reasons and discretely, having no causation or correlation to a more serious issue This incident is expected to occur, as the production line is unable to ensure 100%

efficiency.

When this happens, The machine continues working to finish the batch as normal. Passed parts are proceeded to the next step and failed parts are proceeded to the discard/recycle route after manual inspection.

2. Consecutive trigger: the products failing consecutively or only several units apart. This could have causation or correlation to a more serious issue such as material fault, machine failure, contaminated environment, human error, etc. This incident is not expected to occur and must be prevented, anticipated, and treated.

When a consecutive trigger (defined in the organization as more than 5 products in a given time interval and more than 3 intervals in an operating cycle) is observed by either the operator or the machine's sensor system, the notification system alerts the people in charge for further action, and the machine pauses.

Here, the batch was brought to the technician for checking. However, this activity requires stopping the entire batch including the actual passed product, which generates no added value.

Moreover, despite having a high pass percentage, the production line, when relying on the sensor system, could easily meet type 1 error, which is the false positive, making failed product passing as qualified. This could also happen when the product failed consecutively but the machine could not detect it, leading to damage and/or waste of the entire batch.

		Reality	
		True	False
Measured or Perceived	True	Correct 😊	Type 1 error False Positive
	False	Type 2 error False Negative	Correct 😊

Figure 3.1 Measured vs. Reality matrix

This error could heavily affect the manufacturer, customers, and end users, impacting revenue, manufacturing quality, and reputation. Hence it is crucial to detect and decrease this type of error.

3.2.2. Approach

For this case study, the project team creates randomized trials, taking batches to check manually and find a failure pattern in order to draw a decision that helps detect and decrease false pass error.

Originally, the engineers speculated that there could be 4 detection strategies:

1. First 3 passed products
2. Last 3 passed products
3. First 3 failed products
4. Last 3 failed products

For the mentioned strategies, the technician will audit the 4 closest units to the product



Figure 3.2 Quality Audit scheme

The audit aligns with the pre-existing audit scheme normally implemented. Hence, the technicians are able to perform their acquainted activity with ease

However, during the case study, the engineers noticed another audit pattern, which is checking all failed units, speculating that this could make detecting consecutive failures easier. For this strategy, the technician will audit the 2 closest units to the product

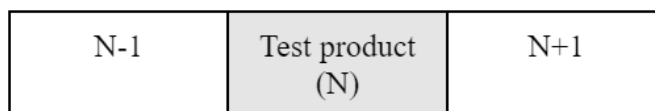


Figure 3.3 All fail audit scheme

We exclude the N+-2 due to the fact of time delays for big unit size. Also, we focus on the batch fail instead of all unit sampling to reduce the waiting time.

Case	Operator	Machine	Shift
1	1	1	1
2	2	1	2
3	1	1	2
4	1	2	2
5	2	1	1
6	2	2	1
7	1	2	1
8	2	2	2

Table 3.1 Design of Experiment

Here, we shared 8 cases with 5 instances per case for speculating influential factors. By means of analysis, the potential impact of the circulating environments in the model of 5M+E are taken into account. Those analyses will drive the line structure of manufacturing including production supervisors, quality officers and engineers to make sure the performance of shifts is appropriate for resource distribution.

Specifically:

- Operator: 2 Operators who are responsible for the working of the machine, Operator #1 and Operator #2 (anonymized). Due to the different levels of skills and experience, personality and work styles are reviewed to do the commonality analysis to acknowledge the quality of the batch. Moreover, the operators' shifts and machines also are factors that play out in the control of the quality of the products. Hence, the instances of the cases are picked randomly in those qualified occurrences.

In reality, there are many operators in the company. But for this pilot project, only 2 were picked randomly.

- Machine: in this project, 2 lines of manufacturing in the factory producing the same products is considered to conduct the case study of machine #1 and machine #2. Because there may be a difference in OEE of the machines, resulting from the difference in performance, machine's durability, and age. They are picked randomly to categorize the difference.

In reality, there are many machines in the company. But for this pilot project, only 2 were picked randomly.

- Shift: Checking the difference in shift's performance, and that is of the fact shift 1 is from 7 am to 7 pm, which enhances the productivity with the support from engineer teams and maintenance teams while shift 2 is under the manufacturing team's supervision alone led by shift leaders. All those factors play a significant role in the quality of care.

After that, the data scientist worked on the data tables and produced a summary analysis of the study before presenting it to the stakeholders.

Based on the result of the case study, the board could agree on a plan to deal with the product quality improvement.

3.2.3. Proposed impact

The main impact would be defect detection, handling, and possible reduction to the manufacturing operation. The additional impacts such as decreased waste in transport, inventory (less space and transportation required to store defective items), human motions, waiting (faster detection and reaction to defect), overproduction, and overprocessing were also feasible and beneficial.

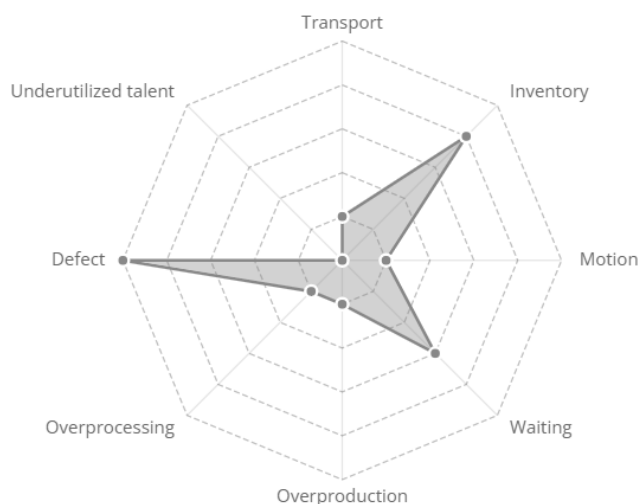


Figure 3.4. Proposed impact

3.3. Method for data collection

3.3.1. Dataset scope

The products under consideration for this study were identified by product code. The research dataset contains 40000 units, categorized with 2 machines, 2 shifts, and 2 different operators, retrieved from the machine and manual sources during the manufacturing process. The collected dataset also contained categorical data to account for variations across product sub-families. Over the course of two months, information needed for this research was acquired. The time period more than a year was used to account for seasonal fluctuation, changes in arriving raw materials, the influence of technician rotating shifts, vacations, and other frequent process and tool variations. It is reasonable to assume that the supplied dataset is trustworthy and indicative of the selected product as a consequence.

3.3.2. Data sources

The analysis of the pass/fail results of units and the relevant product information were the primary data sources. The analyses were carried out by high-precision machines and trained manufacturing technicians who were experts in the relevant tests. The product's final grade and failure mode were included in the analysis. For the simplicity of the project, they were classified into 2 main passed & failed modes. Every unit produced generates output, which is saved in the tablet press SQL database. The data were then uploaded to a server in real-time, where the data scientist can retrieve, process, visualize, or extract the data. This data is semi-structured and needs to be cleaned and organized before it can be used.

The database program used in the enterprise is MongoDB. MongoDB is the most widely used NoSQL document-oriented database. The document model is ideal for unstructured data because it allows users to effortlessly mix and arrange information from numerous sources.

3.3.3. Data collection methods

Product: Data were automatically retrieved for each product using RFID (Radio-Frequency Identification)

Pass/Fail: The machine automatically detects effects and the specific kind of defect using its sensor system. For the simplification of this research, the specific reason is stored on the overall dashboard only and omitted in the individual report.

Actual Pass/fail: The manufacturing technician performs a quality check on individual products

Proof of failure: Though not actually presented in the overall finalized data and presentations, the picture of defects was captured by the technicians to store for reporting and potential quality improvement projects

3.4. Method for data processing

3.4.1. Machine-generated data processing

First the data is extracted from the MongoDB database using SQL code or SQL based software and the joint in the data tables in the form of CSV. Afterwards, the data is put into an excel workbook for further processing.

3.4.2. Human-generated data processing

The data, after being input from the technicians, is stored on the database, transformed into CSV by the author. From then, the author added the actual result into the excel table generated from the previous processing.

3.4.3. Statistics processing

Overall, the below formulas are utilized to calculate the yield and top loss of the process in the whole batch(which helps us to narrow down the tacked target to maximize the efficiency). What we are aiming is to improve and target the confidence interval to be around 95% and get triggered accurately when excursion appears.

- Yield %

$$\frac{\text{Number of passed units}}{\text{Total number of units}}$$

- Actual Yield %

$$\frac{\text{Number of actual passed units}}{\text{Total number of units}}$$

- Minimum sample need to take to detect 1 error

$$n = N \times \frac{\frac{Z^2 \times p \times (1-p)}{\alpha^2}}{N-1 + \frac{Z^2 \times p \times (1-p)}{\alpha^2}}$$

With:

- n: sample size to detect 1 error
- N: batch size
- Z: Z score
- p: defect percentage = 1 - yield percentage
- α : alpha level = 1 - confidence interval
- Confidence interval: 95%

Average yield percentage and average actual yield percentage (listed above) were also specifically used for the DoE to calculate influential factors that could contribute to the defect rate. In the below tables, the yield percentage is noted as Y

The regression equation can be represented as

$$y = p_0 + p_A x_A + p_B x_B + p_C x_C + p_{AB} x_A x_B + p_{BC} x_B x_C + p_{CA} x_C x_A + p_{ABC} x_A x_B x_C$$

The p's are the parameters that are determined by using the outcome matrix

CASE	OPERATOR (A)	MACHINE (B)	SHIFT (C)	Y	AB	AC	BC	ABC
1	1	1	1	0.9234	1	1	1	1
7	1	-1	1	0.9222	-1	1	-1	-1
3	1	1	-1	0.9266	1	-1	-1	-1
4	1	-1	-1	0.9214	-1	-1	1	1
5	-1	1	1	0.9180	-1	-1	1	-1
6	-1	-1	1	0.9324	1	-1	-1	1
2	-1	1	-1	0.9222	-1	1	-1	1
8	-1	-1	-1	0.9260	1	1	1	-1

Table 3.2. Factorial design - Machine

CASE	OPERATOR (A)	MACHINE (B)	SHIFT (C)	Y	AB	AC	BC	ABC
1	1	1	1	0.8964	1	1	1	1
7	1	-1	1	0.8988	-1	1	-1	-1
3	1	1	-1	0.8990	1	-1	-1	-1
4	1	-1	-1	0.8978	-1	-1	1	1
5	-1	1	1	0.8942	-1	-1	1	-1
6	-1	-1	1	0.9104	1	-1	-1	1
2	-1	1	-1	0.8912	-1	1	-1	1
8	-1	-1	-1	0.9012	1	1	1	-1

Table 3.3. Factorial design - Actual

3.5. Method for data visualization

3.5.1. Power BI

Power BI makes it easy to compare and explore data both on-premises and in the cloud, enabling easy and safe collaboration and sharing of customizable dashboards and live reports with colleagues and companies.

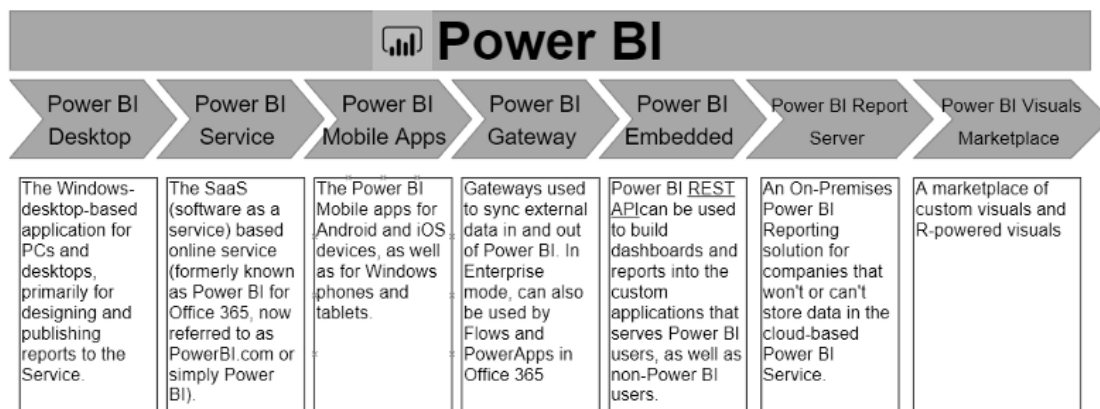


Figure 3.5 Power BI environment

Power BI has several advantages that set it apart from other analytical tools. It offers both software and a cloud-based user interface. Data warehousing, data exploration, and interactive dashboards are all available. Custom visualizations can be loaded, and the system is highly adaptable across the entire enterprise.

For this specific research project, Power BI Desktop and Power BI Service were used to create real-time dashboards in the shift for shift leaders, supervisors, and engineer team to track and report the production.

Power BI dashboard illustrates charts and diagrams to monitor overall quality and top loss for future decision making process.

The overall quality will be illustrated with a barchart including a constant line at 95% (ideal yield) according to time and the reproducibility of the gauge for error configuration. The count of the failure will be placed in the pareto charts and pie charts per machine, to make decisions on the most occurring ones.

3.5.2. Microsoft Excel

Excel has long been a preferred tool for conducting data analysis due to its simplicity and flexibility. It provides great tools to improvise correctly with the data collected from the shop floor (mostly CSV) in the format of tables in spreadsheets. It improves the ways of decision making by sharing the significant data tools in the data tab on the toolbar.

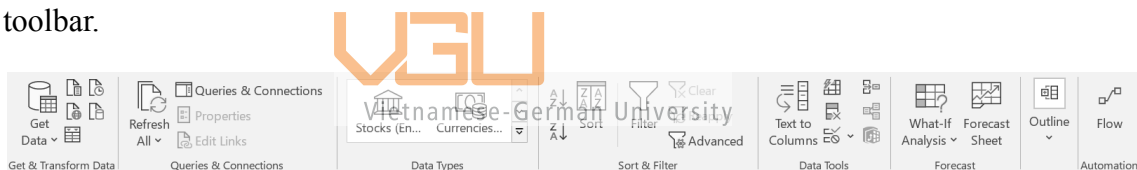


Figure 3.6. Several tool for data analysis in Excel

In the individual report, all the relevant data of the batch is displayed as a table, consisting of number (1000 units), machine code, operator, shift product code, pass/fail, real pass/fail. Although a lot of data was processed and was present in the database, only the values relevant to the DDDM process are shown.

In the summary report, all cases have their yield, actual yield, and detection method listed as a table. The author also created an additional table for the final calculation to assist with DDDM.

3.5.3. Microsoft PowerPoint

For many years, PowerPoint has been utilized for various professional meetings and data visualization. They are widely used as they allow a large volume of data to be displayed on one screen and are simple to use. PowerPoint has been updated several

times throughout the years, increasing the number of capabilities available to users. These advancements have enabled the presentation of photographs, videos, and now even websites in PowerPoint.

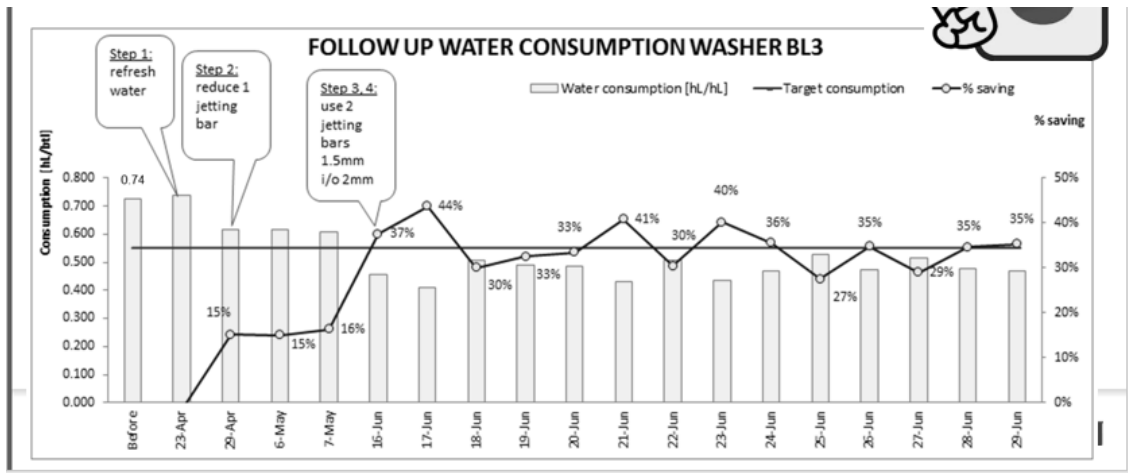


Figure 3.7 Illustration of data chart presentation in Powerpoint

Powerpoint reports the batch file and related data throughout the indicators for alerts by notice in the batch. It also shows the formula and result of yield and actual sample size needed to get the proper error. Also, when it comes to pivot files, charts and tables are in use.



Vietnamese-German University

4. Application in a worker planning project

The author conducted this project in company #2. Here, big data is implemented in a project to support the analysis and planning of the organization's workforce.

4.1. Project Strategy

The project strategy guides the effort to produce the desired strategic results. The project will use these high-achieving concepts to decide what to do and how to do it in order to follow the legal, financial, economic, and functional strategies of the organization. Existing frameworks and models might help the leader and the team define their project approach.

4.1.1. Waterfall project management

A project's linear continuity from beginning to end is emphasized by the waterfall approach to project management. This front-loaded approach, which depends on rigorous preparation, thorough report, and consistent execution, is commonly used by engineers.

The Waterfall method, for example, in a software development project, usually contains the following steps:

1. Requirements
2. Design
3. Implement
4. Test
5. Deploy and maintain

The method is a tried-and-true framework for management that is clearly defined and simple. Each contributor can efficiently arrange their time throughout the project since they are aware of exactly what must be accomplished and when it must be completed because the goals are often laid out from the beginning.

Its style, in contrast, is less flexible or agile as the game progresses since it places a premium on meticulous project preparation and adherence to predetermined deadlines. Later-stage changes might be time-consuming, challenging, and costly.



4.1.2. Agile project management

Large projects are divided into smaller, easier to administrate pieces that are completed in brief iterations across the various phases of the project's life cycle using the agile methodology, an incremental project management strategy. Agile businesses are better equipped to do jobs more rapidly, adapt to changing project requirements, and streamline their processes.

Agile aims to provide quicker iterations and much more regular product launches. Project teams are prepared to react to shifts in the customer demands because of the compressed time period.

In a software development project, for example, the Agile method usually includes the following steps:

1. Plan
2. Create roadmap
3. Release planning
4. Plan sprint
5. Hold meetings
6. Review sprint



Vietnamese-German University

Agile is a more flexible, iterative approach that addresses preparing, styling, implementation, and checking tasks in relatively short, repeatable iterations, in contrast to the rigid Waterfall methodology, which necessitates thorough preparation in advance and mandates that each phase be finished prior to actually moving on to the next.

The consumer often plays no active part in Waterfall after their needs have been identified. Agile gives the client more influence over the development process. Instead of milestones, Agile uses "sprints," usually last between 7-14 days.

Hence, for this specific project, which requires a web application with fast-changing data and user-orientation, the agile methodology is implemented.

4.1.3. Project timeline

The project timeline began with the Product Owner and business analyst working with stakeholders to collect the voices of customers to Critical-to-quality (CTQ), etc for decision making on user interaction, required data & data source. The product owner then continuously generates a backlog, while the business analyst works with the

shopfloor to create preferred visualization, and predictive/forecast methods. The next step is to run the project by sprints, spending around 2 weeks/ sprint, with the inclusion of continuous testing and feedback cycles by stakeholders and the team. Finally, the application is set to roll out after finishing the beta test. Even after that, the team continuously collects feedback and improves the apps with further sprints.

	Month 0	Month 1				Month 2				Month 3				Month 4
		Week 1	Week 2	Week 3	Week 4	Week 1	Week 2	Week 3	Week 4	Week 1	Week 2	Week 3	Week 4	Week 1
Work with stakeholders and develop project charter	█													
UX/UI Design		█	█											
Database connection		█	█											
Sprint 1: Select worker by skill				█	█									
Sprint 2: Assign training						█	█							
Sprint 3: Upload/Download multimedia files								█	█					
Sprint 4: Edit misc. information										█	█			
Pilot testing				█	█	█	█	█	█	█	█	█	█	█
Project rollout														●

Table 4.1. Project timeline

4.2. Conceptual design

While current HR data analytics is able to handle worker’s details. The use of various third-party, in-house programs, or extensive data tables could be confusing and insecure. Hence, the Digitalization and Technology team has proposed the all-in-one application for worker control, headcount planning, training, and development.



4.2.1. Technical Design

This is the overall design for the development team

The developer team worked with database, cloud service, front end (including UX/UI design), and back end.

The cloud service used by the corporation is Azure.

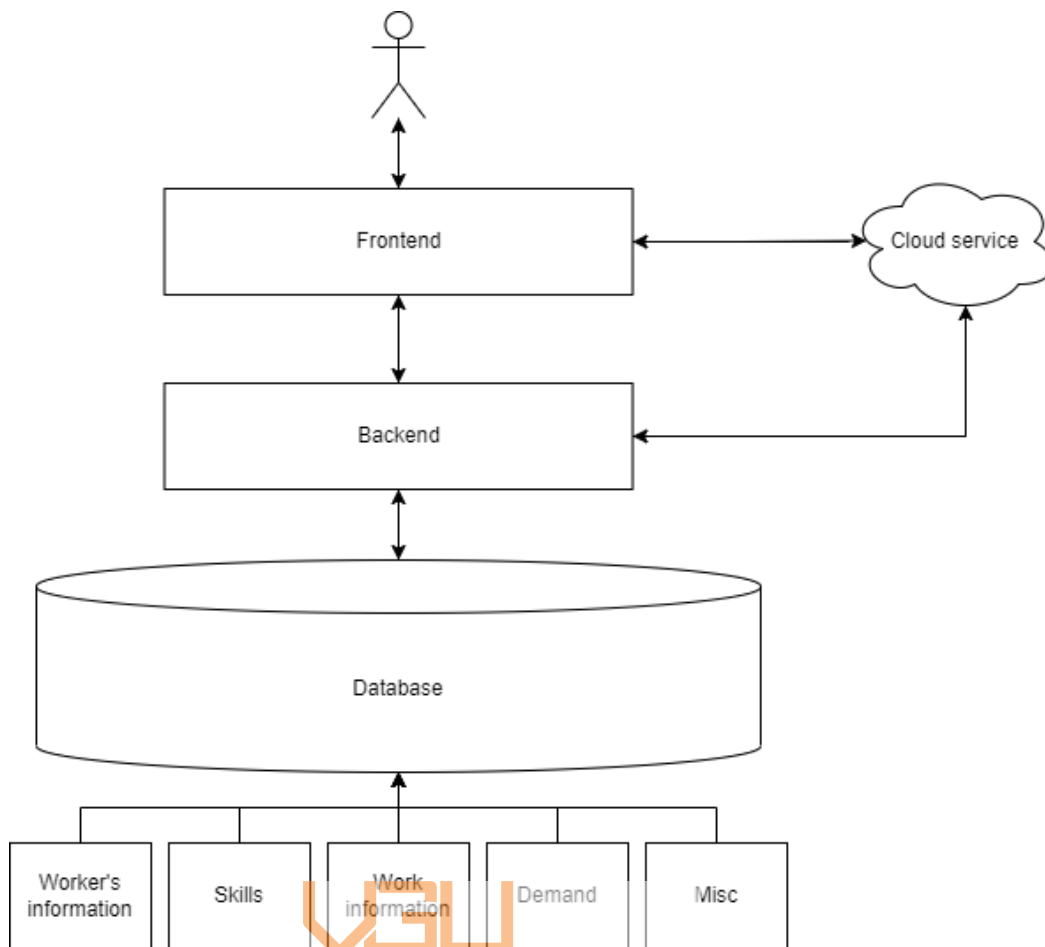
Database: PostgreSQL connected with DBeaver

Backend: Java Springboot in IntelliJ IDEA application by JetBrains

Front end: javascript with angularJS framework

Additionally, to support the development of the application, the author also used Visual Studio Code and QuantUX programs

The official UX/UI was designed by the UX designer



Vietnamese-German University
Figure 4.1. Technical usage

4.2.2. User experience

Below is the conceptual design for the end-user

End users include but are not limited to HR (Learning & Development, HR Business Partner, HR Service, Compensation & Benefits), shop floor leaders (team leader, supervisors, line managers, factory managers), support team (production planner, headcount planner, logistics, outsourced HR & Talent Acquisition)

The User interface is presented as a single web application and is uploaded on the company's shared web application resource. The key user will access the web, sign in, and perform actions on the application.

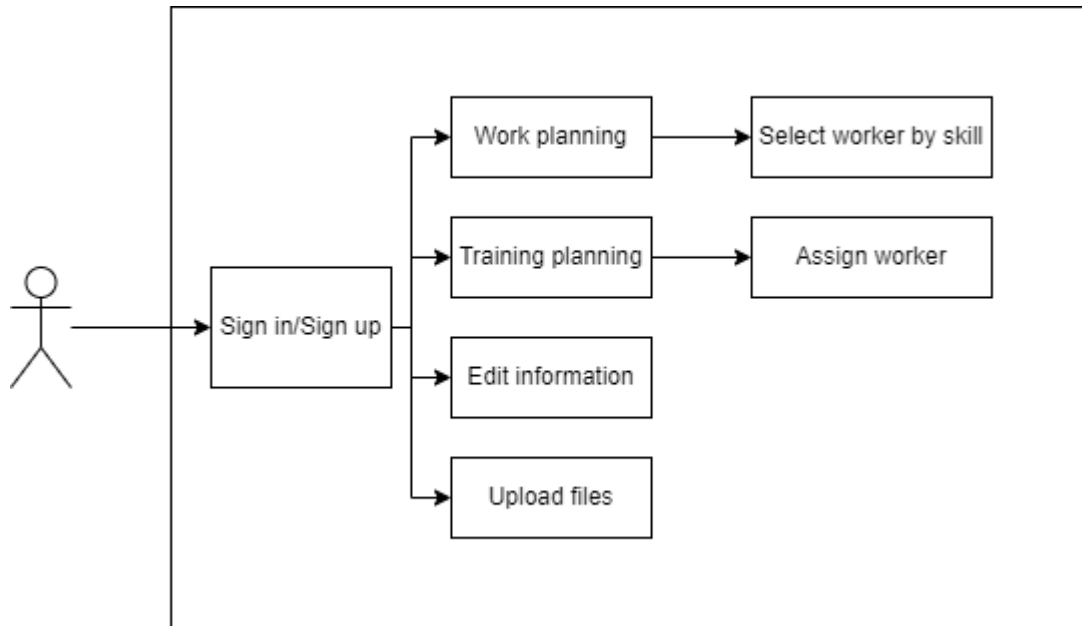


Figure 4.2. End-user interaction

4.2.3. Proposed Impact

The main impact would be the increase in workforce skill and experience, utilizing the human capability in the factory. The additional impacts such as a decrease in defect rate, inventory (less space required to store defective items), human motions, and overprocessing were also feasible and beneficial as a side result of improved workers.

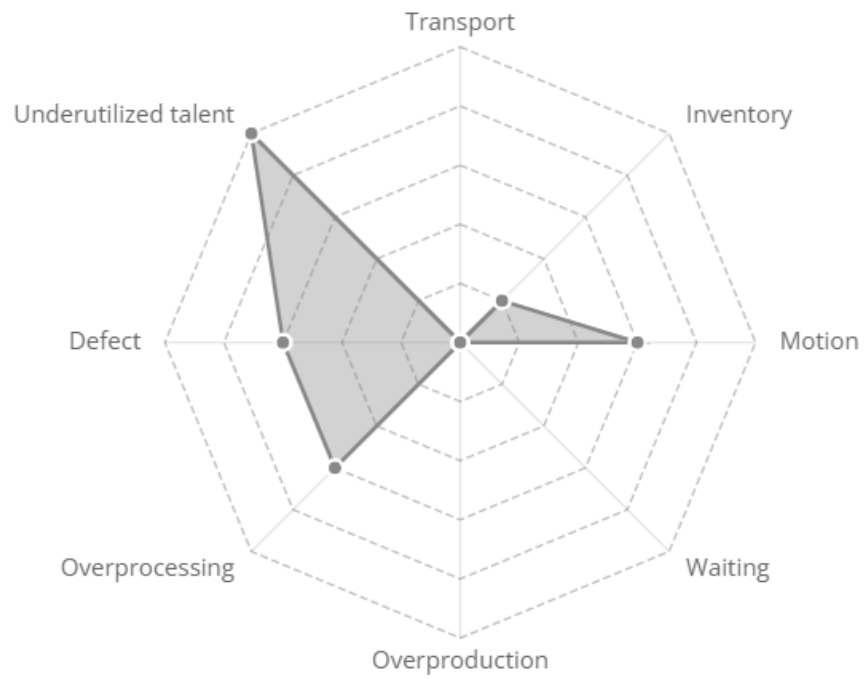


Figure 4.3. Proposed impact
 Vietnamese-German University

4.3. Method for data collection

The data source is varied depending on the nature of the data.

- Worker’s information (name, ID, education, phone, etc.): Provided by the worker in their resume (Physical papers and e-forms)
- Worker’s skill: specified by the worker’s original output in their resume, their work-related skills, and their on-the-job training, the data is collected by the HR person responsible for learning and development after each training session. (Physical papers & data retrieved from internal online training website)
- Work information: Collected based on primary info (Position, level, supervisor, workstation, work center) and secondary information (Production planning, OT & Time off, substitute, future consideration (Stored on Database))
- Demand: Collected from production planning & finance department (Excel & Powerpoint presentation)
- Worker’s attendance: this includes time in, time out, leave days, and overtime. Though this has been applied in another outsourced application, the data is still collected for assessment purposes.
- Other data depends on special inquiries. This includes but does not limit to specific queries during career, promotion offers, performance assessment, etc.

For the confidential nature of the data, employee’s personal data is anonymized in the thesis, and the study will focus on the big data application aspect

Skill:

Data name	Description	Value type
No	Number	Numeric (int)
Type_skill_code	Skill group’s code	Numeric (int)
Type_skill	Skill group’s name	Varchar (255)
skill_code	Skill’s code	Numeric (int)
skill	Skill’s name	Varchar (255)
level	Skill’s level (from 1 to 5)	Numeric (int)

Table 4.2. Skill data table

Worker:

Data name	Description	Value type
No	Number	Numeric (int)
Employee_code	Employee's ID	Numeric (int)
Last_Middle_Name	Employee's last and middle name	Varchar (255)
First_Name	Employee's first name	Varchar (255)
Full_Name	Employee's full name	Varchar (255)
Current_position	Employee's position	Varchar (255)
Current_workstation	Employee's workstation	Varchar (255)
Current_supervisor	Employee's supervisor	Varchar (255)
Current_workcenter	Employee's workcenter	Numeric (int)
Contract_date	Employee's contractor date	Date
Contract_date_expiry	Employee's expired date	Date
Contract_expired	Employee's contract status	Boolean
Onsite	Employee's on site status	Boolean
Remark	Additional note	Varchar (255)

Table 4.3. Worker data table

4.4. Method for data processing

4.4.1. Data pre-processing

- Physical papers: Transformed into data on the database by the person in charge. The physical files were nevertheless scanned and uploaded to the corporate's internal cloud database and have their path stored as part of the DB to keep as proof for possible audit
- E-forms: usually generated by google forms or intra-organizational form/survey, were directly converted to excel files, pdf, or images for HR and managers. Necessary information was sorted out and input into the database
- Excel files and PowerPoint presentations are processed similarly to E-forms
- Much of the other data has already been processed and stored on the database beforehand

All data was compiled and stored on the database, stored on PostgreSQL, and administered with DBeaver, both open-sourced but secured database management tools.

4.4.2. Data processing tools

This project used open-source coding platforms and tools, specifically:

- Visual Studio Code with Spring Boot and Angular frameworks
 - An easy-to-use code editor with task management, version management, and troubleshooting capabilities is called Visual Studio Code.
 - A Java-based framework for creating microservices is called Spring Boot. It is used to produce standalone, production-ready spring apps.
 - Based on HTML and TypeScript, Angular is a framework and architecture for creating applications.
- JHipster: a web application and microservice architecture development platform that allows the developer to quickly create, build, and deploy modern online applications and microservice designs.
- Apache Maven: a software project administration and comprehension platform.

For the coding language, the following languages were used:

- Java: In addition to huge data analysis and integrated devices, it is used to create desktop and mobile apps. According to Oracle, the corporation that owns Java, it is considered one of the most widely used language
- Typescript: Typescript translates to JavaScript and is designed for the development of large apps.
- XML, Caddyfile, and several coding languages were also used in various small elements to assist with the web application.

For the processing, the detailed codes were too long and complicated for the scope of this bachelor thesis, so the following general data processing steps were taken, Divided into Back-end and Front-end.

4.4.3. Back-end processing

Most of the data was handled in the backend.

- User Account and Authentication (UAA): UAA's main function is to serve as an OAuth2 provider, providing tokens to client apps that work on behalf of Azure Cloud users. UAA may authenticate users using their Azure Cloud credentials and act as an SSO service utilizing those or other credentials in conjunction with the login server.
 - The database for UAA is stored on the organization's internal UAA server.
 - Additionally, There was also a manual sign-in (separated from the Azure Cloud), in which the user account and authorization are stored in the database.
- Data tables: The data were directly pulled from the database to the backend through SQL query connection,
 - The column names were edited for a user-friendly interface
 - SQL data types were changed to respective java data type
 - For the relational database, the tables were joined with their respective connection: one-to-one, one-to-many, many-to-one, etc.
- Table repository: JPA-based repositories, which are utilized under the JAVA spring framework, are frequently referred to as Java repositories. Repositories define a novel, straightforward approach for storing, modifying, and extracting information from backend JAVA applications.

- In this project, the repositories were mostly used to support the search function
- Data Transfer Objects: in this project, Data Transfer Objects were used to transport data between tasks to decrease the number of method calls. This is to support the front end in retrieving information from the database.
- Mappers: the mappers were used to connect the data transfer object with the corresponding tables and elements in the database.
- Services: serving as the main interaction between the frontend and the database, this handled information editing, file uploading, filtering, suggesting, and updating.

Several criteria were provided for ease of suggestion. During the rollout however, the team continued to receive feedback and improve the algorithm.

Indicator (generalized)	Suggestion (generalized)	Processing action (generalized)
Employee taking an abnormal amount of leaves, or take leaves consecutively without clear reason	Notify direct manager to check up on the employee Send notification to the employee and manager through email	Analyze work performance, workload, attendance, and other data to find abnormality
Employee voluntarily taking up training and courses	Notify direct manager for mentorship and work advice Suggest aligned courses Notify manager for recognition	Analyze course pattern to suggest courses and career path
Employee's performance decreasing	Notify employee Notify manager to check up on the employee	Analyze work performance, workload, attendance, surveys, and other data to find pattern and/or

		abnormality
Employee's contract duration ending	Notify manager and employee for next step Suggest next step	Analyze current work scope, opening positions, performance, completed training, etc. to suggest between letting go, extend contract, or promote to new position
Employee not taking any training for a long time	Notify direct manager for mentorship and work advice Suggest new or refresh courses	Analyze current workload, performance, and other data
Employee demanding for a raise or promotion	Notify employee and manager for communication, review, and suggested raise or position Vietnamese-German University	Analyze work performance, workload, attendance, current position, current salary, open positions, salary range, surveys, and other data
Employee taking up too much overtime	Notify direct manager for mentorship and work advice Suggest wellness check	Analyze work performance, workload, attendance, and other data
Employee's performance increasing	Notify manager for recognition Suggest aligned training	Analyze work performance, workload, attendance, and other data
Employee showing signs of leaving	Notify direct manager for communication and possible procedures	Analyze work pattern, work performance, workload, attendance, surveys, and other data

Employees showing sign of work overload	Notify people in charge for communication and planning	Analyze work pattern, work performance, workload, attendance, surveys, and other data
---	--	---

Table 4.4. Data processing and action

4.4.4. Front-end processing

Front-end processing mostly handle the User Experience/User Interface and information exchange

- User Account and Authentication: When the user use Single Sign-On or manual sign-in, the front-end process the input and send the token to the user database for authentication
- Filtering by criteria: Using sets of interactive filters to utilizing the usability of the application
- The access and interaction levels of user synchronize with their roles in the company, which leads to the suggestion filtered with their levels (for example: shift leader comes with OT schedule, assignments, manager with KPI and OKR, and HR with promotion/ salary raised records and databases)

4.5. Method for data visualization

For visualization, as mentioned in the previous section, The final goal of this project is to create an all-in-one application for the HR and management team to manage and utilize the workforce.

Hence, the data is presented in a style of a web application, showing information relevant to the user's demand. The UX/UI is designed by the UX Designer of the corporation. Here, for the sake of disclosure, the author is using QuantUX, an open-source prototyping tool to replicate the interface of the application. The actual interface is sent separately to the thesis supervisor.

4.5.1. Worker selection by skill

The visualization is presented for users as a filtering objective: first, we manage to choose position, and then skill group. Afterwards, the complete skill matrices are shown for proper application.

The interface consists of three main columns and a 'Next' button:

- Pick a position:** Radio buttons for Machine Operator (selected), Machine Technician, Supervisor, and Engineer.
- Pick a skill group:** Radio buttons for EHS (selected), General skill, Technical skill, and Special skill.
- Choose skill(s):** Checkboxes for Electrical safety, Workplace Protection, Shopfloor Ergonomics, Workplace Ergonomics, Working in confined place (checked), and Working at height.

At the bottom center is the logo for Vietnamese-German University (VGU).

Figure 4.4. Skill selection panel

4.5.2. Training assignment

There are 2 ways to access the data on this assignment, through the tab training to assign the worker with the right skillset and the level. After that, the list of workers will appear in tables to be selected. The other way around is to get to the worker's page and select the worker's name on the general site and find the suggested training.

The interface includes a form with the following elements:

- A dropdown menu for skill group, currently showing 'EHS'.
- A dropdown menu for skill, currently showing 'Shopfloor Ergonomics'.
- A 'Confirm' button.
- A dropdown menu for 'Choose shift'.
- A dropdown menu for 'Choose position'.

Below the form is a table with the following data:

No	Employee_Code	Last_Middle_Name	First_Name	Current_Position	Current_Segment	Supervisor	Shift
1	3733171	Le Thi Hoai	Lam	Machine technician	KRO276A	Do Tri Nguyen	1
2	3782100	Tran Viet	Hoang	Machine technician	DGB144A	Nguyen Hoang Duc	1

Figure 4.5. Training tab

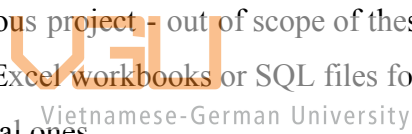
The data assess the training data for workers, mostly to evaluate the efficiency of the last training session and count the number of untrained workers for arranging future training. To the potential worker for promotion, the training list would be suggested, and can be printed to HR.

4.5.3. Edit information

The data system announces the needs of editing (i.e: current position, future position, supervisor, line manager) and corrects the information misplaced or wrong. Not only that but the system also follows the whole process performance of the employees (learning capacity, working OKR, KPI, etc) for action on the integration plan.

4.5.4. Upload/Download resources

The system has an upload, download, print document, resource feature to accommodate the best UX/UI for HR and shop floor' supervisors. It can also use the outsource application (in the previous project - out of scope of thesis and confidential) to transfer the file tables (CSV) to Excel workbooks or SQL files for storing in databases or taking the databases into physical ones.



5. Result

5.1. Production Quality Management

Dashboard following up the process performance through power BI:

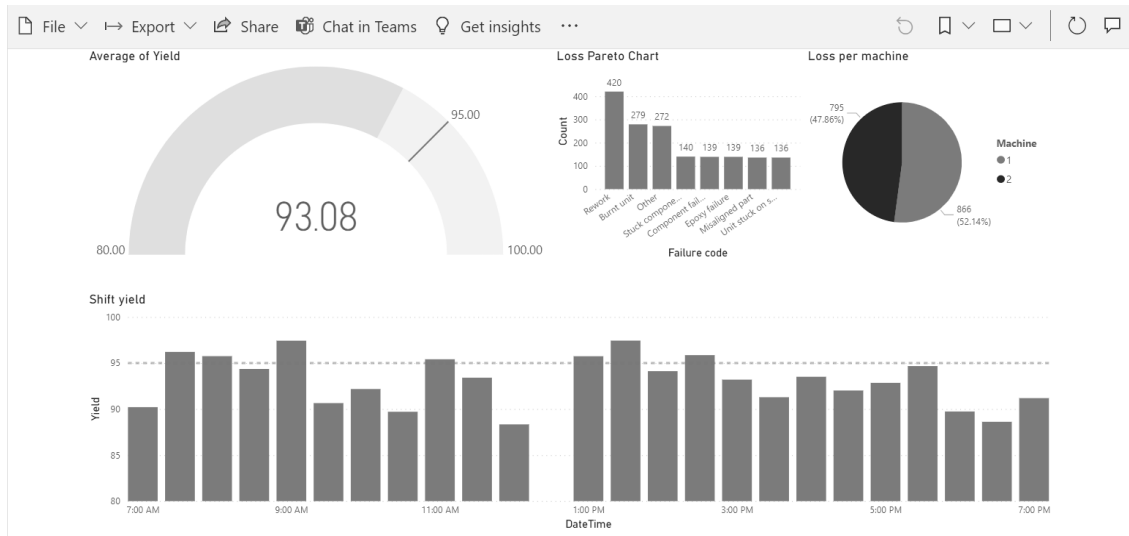


Figure 5.1. Generated overall quality dashboard

The result demonstrates that the shift yield is currently around 93.08 %, lower than the ideal value (95 %), which shows the variation of quality in the batch needs to be investigated further. Certain methodologies will be in use, but the most common is commonality analysis, to assess the root cause for better understanding and take immediate action. The most usual failures and top losses, such as: rework, burn are instances, which after certain actions, FMEA are made to make sure that these are under control, and preventive actions are integrated.

Specifically, for our case study:

Mean yield	Actual mean yield	First 3 passed	Last 3 passed	First 3 failed	Last 3 failed	All failed
%		15.00%	2.50%	25.00%	22.50%	92.50%
Min sample size		84	95	78	81	78
Max sample size		108	95	119	115	119
Average sample size		92	95	100	97	98

Table 5.1. Final success rate table

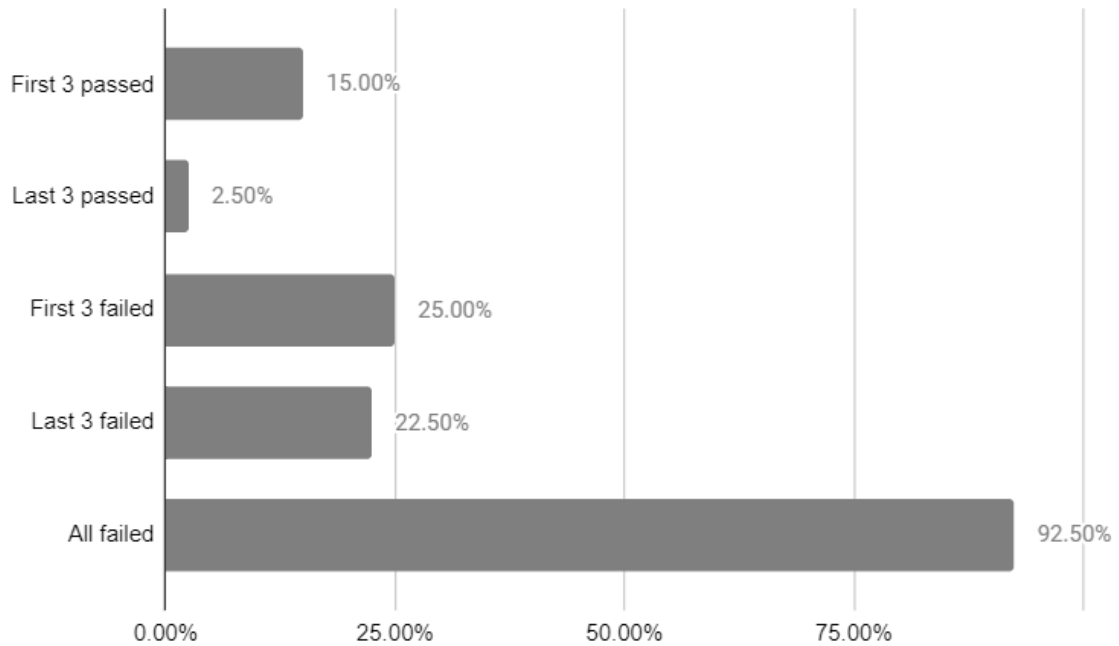


Figure 5.2. Comparison of case study success rate

The influence of the factors is taken into consideration.

	A	B	C	AB	AC	BC
1	0.8980	0.8952	0.9000	0.9018	0.8969	0.8974
-1	0.8993	0.9021	0.8973	0.8955	0.9004	0.8999
Δ	-0.0012	-0.0069	0.0027	0.0063	-0.0034	-0.0024
$\Delta/2$	-0.000625	-0.003425	0.001325	0.003125	-0.001725	-0.001225

Table 5.2. Influence table

Here, the value suggested weak influence from the considering factors on the yield, with delta all less than 1%. However, the machine, and the combination of operators working on the machine have slightly higher influential value on the yield compared to other factors. This suggests that the operator can have different working agility depending on the machine, or that the root cause could be other factors, i.e material, thermal and electrical failure, oxidation, etc. And there could be another case study needed to deep dive in these matter

For the final decision, the all failed detection tests for failure and false pass error provides the best yields for quality inspection. However, the time and efforts for such labors is arduous, which led us to the second option is to detect the first 3 failed, but it hardly satisfied us with the low percent of success (25%).

5.2. Worker Planning

Pick a position

Machine Operator

Machine Technician

Supervisor

Engineer

Pick a skill group

EHS

General skill

Technical skill

Special skill

Choose skill(s)

Electrical safety

Workplace Protection

Shopfloor Ergonomics

Workplace Ergonomics

Working in confined place

Working at height

Next

Figure 5.3. Skill selection panel



Vietnamese-German University

EHS

Shopfloor Ergonomics

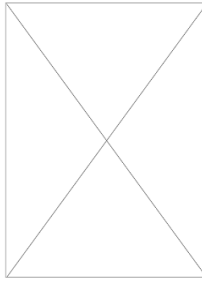
Confirm

Choose shift

Choose position

No	Employee_Code	Last_Middle_Name	First_Name	Current_Position	Current_Segment	Supervisor	Shift
1	3733171	Le Thi Hoai	Lam	Machine technician	KRO276A	Do Tri Nguyen	1
2	3782100	Tran Viet	Hoang	Machine technician	DGB144A	Nguyen Hoang Duc	1

Figure 5.4. Training tab



Full name: Le Thi Hoai Lam
 Employee ID: 3733171
 Current position: Machine Technician
 Current supervisor: Do Tri Nguyen
 Shift: 1
 Onsite: Yes

+ The employee has not taken any training in a long time. Consider assigning them new or refresh courses

Employee information		Skill information			Performance information			
SkillGroup_ID	Skill_ID	SkillGroup	Skill	Level	Date of certification	Remark	Attachment	
131	13101	EHS	Electrical safety	3	2021/07/03			Assign Edit Delete
131	13102	EHS	Workplace protection	2	2022/04/01			Assign Edit Delete
131	13103	EHS	Shopfloor ergonomic	0				Assign Edit Delete

Figure 5.5. Worker tab

In summary, this is the prototype for shop floor usage, with the continuous cycle of feedback-improvement with side usage as human resources assessment for higher management.



Vietnamese-German University

6. Discussion

This section discusses the achievement, limits, and possible outlook on big data analytics and its application in the manufacturing industry.

6.1. Big data analytics

Analyzing data becomes increasingly difficult as the amount of data increases. The key problem, in addition to managing a vastly fluctuating volume as well as a variety of data, is extracting meaningful value from it. Analysts and managers must have a thorough understanding of big data applications. Companies must accelerate employment programs and invest in important employee training and education.

6.2. Production quality management project

The case study has pointed out the best way to detect type 1 error during production. However, it appears that the root cause for the defect is still not discovered. A future case study could be performed to deep dive in these matters.

Pros:

- + Defined detection strategy
- + Defined importance of main affecting factors
- + In-house project with ensured security

Cons:

- + Time-consuming
- + Detection strategy requires much effort in routing and planning
- + May require more studies to detect underlying factors and root cause

Businesses are facing significant challenges as a result of the globalization of economies. This is driving the manufacturing sector to undergo its upcoming metamorphosis towards predictive manufacturing. Hence, chances appear for this research to be applied into predictive analysis for quality patterns.

6.3. Worker planning project

The web application is on the way to rollout on the shop floor. The digitalization team is still continuously improving for user friendliness and data optimization.

Pros:

- + In-house application with ensured security
- + Frequent update and catchup with stakeholders
- + Ease of navigation for end user
- + Direct communication with database
- + No funding/expense required for contractor

Cons:

- + Time-consuming
- + Restricted use: only for internal desktop and laptops

For future use by upper management and for quicker access, the option for mobile-friendly app development is also taken into consideration



7. Conclusion

Without a doubt, the age of data is now. Big Data Analytics is not just a viewpoint for all organizational levels. This is particularly relevant to the manufacturing process because of the high capital intensity, the tight time limitations, and the vast amount of data that has already been collected. Data is a valuable resource for any organization. Utilizing high-quality data properly can aid firms in developing more accurate forecasts, assessments, and judgment calls. Big data accessibility, availability, and reliability: There is a sizable amount of historical data dispersed over a number of repositories, as well as new data produced every day by billions of interconnected gadgets or created independently by businesses and individuals. Then, in order to enhance and develop a business, it is required to create more suitable and efficient methods for utilizing such data, to navigate them, and to comprehend the reasons behind their acquisition and their quality. Big Data analytics is a constantly expanding notion. As evidenced by the aforementioned examples, its environment offers excellent opportunities for firms from a variety of industries to thrive with a competitive advantage.



Vietnamese-German University

8. Bibliography

- Enofe, M. O. (2017, May). Data Management in an operational context - core. Retrieved April 7, 2022, from <https://core.ac.uk/download/pdf/289955103.pdf>
- Garvin, D., 1987. Competing on the Eight Dimensions of Quality. [online] Harvard Business Review. Available at: <https://hbr.org/1987/11/competing-on-the-eight-dimensions-of-quality> [Accessed 3 May 2022].
- Katal, A, 2013. Big Data: Issues, Challenges, Tools and Good Practices. Big Data: Issues, Challenges, Tools and Good Practices, 1, 404.
- Otchere, Samuel & Nyamewaa, Ennin & Hammond, Frank. (2022). Big Data Characteristics and Innovation Performance in Ghanaian Manufacturing Firms: The Role of the Big Data Team?. Open Access Library Journal. 09. 1-13. 10.4236/oalib.1108378.
- O'Donovan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. J. (2015). Big data in manufacturing: a systematic mapping study. Journal of Big Data, 2(1). <https://doi.org/10.1186/S40537-015-0028-X>
- James Bell Associates. (2018). Guide to Data-Driven Decision Making: Using Data to Inform Practice and Policy Decisions in Child Welfare Organizations. Washington, DC: Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services.
- Bartels S, Richey J. Workforce Management– Impactingthe Bottom-line 2008.3.
- Higgins J, Cooperstein G and Peterson M of HumanCapital Management Institute (HCMI), Top Five Metrics For Workforce Analytics, 2011.
- Taleb, Ikbal & Serhani, Mohamed & Bouhaddioui, Chafik & Dssouli, Rachida. (2021). Big data quality framework: a holistic approach to continuous quality management. Journal of Big Data. 8. 10.1186/s40537-021-00468-0.
- Wang, Keqin & Tong, Shurong & Eynard, Benoit & Lionel, Roucoules & Nada, Matta. (2007). Application of Data Mining in Manufacturing Quality Data. 2007 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2007. 5382 - 5385. 10.1109/WICOM.2007.1318.

Siriwardana, Chandana. (2016). An Integrated Framework for Worker Planning and Supervision in Construction.

Jay Lee, Edzel Lapira, Behrad Bagheri, Hung-an Kao, Recent advances and trends in predictive manufacturing systems in big data environment, *Manufacturing Letters*, Volume 1, Issue 1, 2013, Pages 38-41, ISSN 2213-8463, <https://doi.org/10.1016/j.mfglet.2013.09.005> (<https://www.sciencedirect.com/science/article/pii/S2213846313000114>)

Amine Belhadi, Karim Zkik, Anass Cherrafi, Sha'ri M. Yusof, Said El fezazi, Understanding Big Data Analytics for Manufacturing Processes: Insights from Literature Review and Multiple Case Studies, *Computers & Industrial Engineering*, Volume 137, 2019, 106099, ISSN 0360-8352, <https://doi.org/10.1016/j.cie.2019.106099>.
(<https://www.sciencedirect.com/science/article/pii/S0360835219305686>)



Vietnamese-German University

9. Appendix

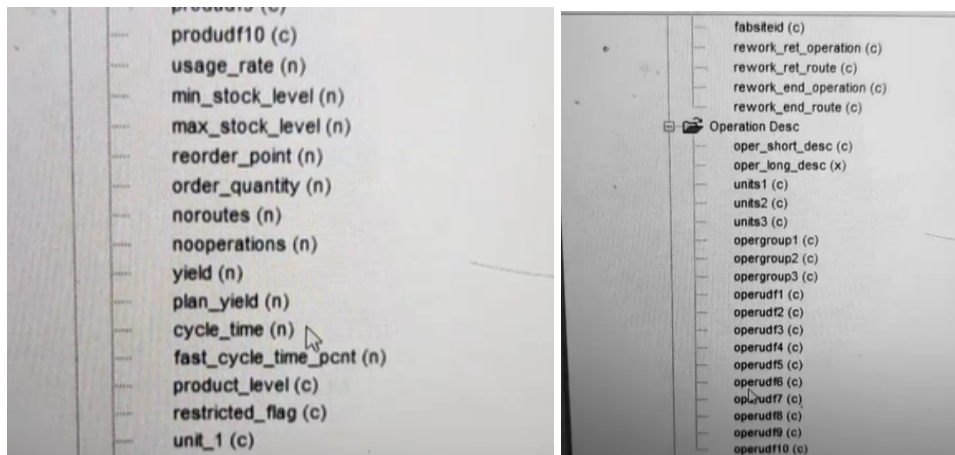


Figure 9.1 Part of the datasource (SQL-based) for the quality management research

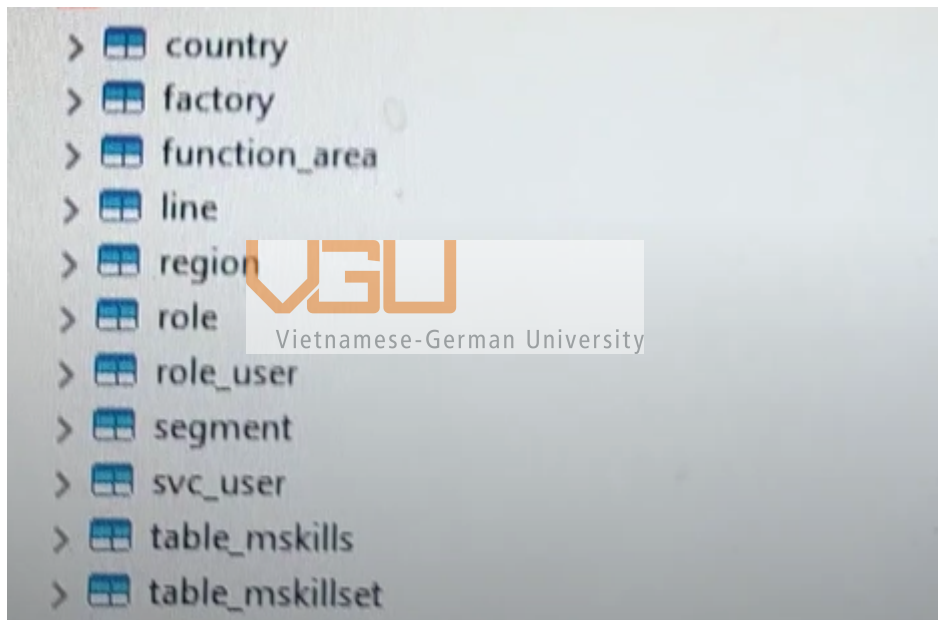


Figure 9.2 Part of the datasource (SQL-based) for the worker planning project

Case	Instance	Yield	Actual yield	Mean yield	Actual mean yield	First 3 passed
1	1	0.930	0.907	0.923	0.896	N
	2	0.923	0.893			N
	3	0.919	0.896			N
	4	0.916	0.881			N
	5	0.929	0.905			N
2	1	0.919	0.888	0.922	0.891	N
	2	0.909	0.882			N
	3	0.924	0.896			N
	4	0.924	0.894			N
	5	0.935	0.896			N
3	1	0.936	0.908	0.927	0.899	N
	2	0.934	0.905			N
	3	0.925	0.895			N
	4	0.907	0.882			N
	5	0.931	0.905			N
4	1	0.931	0.907	0.921	0.898	N
	2	0.931	0.905			N
	3	0.918	0.895			N
	4	0.914	0.888			Y
	5	0.913	0.894			N
5	1	0.914	0.898	0.918	0.894	N
	2	0.912	0.886			N
	3	0.926	0.895			N
	4	0.930	0.907			Y
	5	0.908	0.885			N
6	1	0.942	0.923	0.932	0.910	N
	2	0.930	0.909			N
	3	0.930	0.902			N
	4	0.924	0.903			N
	5	0.936	0.915			Y
7	1	0.920	0.905	0.922	0.899	N
	2	0.912	0.888			N
	3	0.933	0.911			Y
	4	0.927	0.905			Y
	5	0.919	0.885			N
8	1	0.936	0.909	0.926	0.901	Y
	2	0.939	0.914			N
	3	0.924	0.898			N
	4	0.928	0.910			N
	5	0.903	0.875			N
				%		15.00%
				Min sample size		84
				Max sample size		108
				Average sample size		92

Last 3 passed	First 3 failed	Last 3 failed	All failed	Sample size to detect 1 error
N	N	Y	Y	91
N	N	N	Y	99
N	N	N	Y	103
N	N	N	Y	106
N	N	N	Y	92
N	N	N	Y	103
N	Y	N	Y	113
N	N	N	Y	97
N	N	N	Y	97
N	N	N	Y	85
N	N	N	Y	84
N	Y	Y	Y	87
N	N	N	Y	96
N	Y	Y	Y	115
N	N	N	Y	90
N	N	N	Y	90
N	Y	N	Y	90
N	N	N	Y	104
N	Y	N	Y	108
N	N	N	Y	109
N	N	N	Y	108
N	Y	N	Y	110
Y	Y	N	Y	95
N	N	N	Y	91
N	N	Y	Y	114
N	Y	N	Y	78
N	N	N	N	91
N	N	N	Y	91
N	N	Y	Y	97
N	Y	N	Y	84
N	N	N	N	102
N	N	Y	Y	110
N	N	N	Y	88
N	N	Y	Y	94
N	N	N	Y	103
N	N	Y	N	84
N	N	Y	Y	81
N	N	N	Y	97
N	N	N	Y	93
N	Y	N	Y	119
2.50%	25.00%	22.50%	92.50%	
95	78	81	78	
95	119	115	119	
95	100	97	98	