Supply Chain Efficiency Prediction: Leveraging Machine Learning for Improved Accuracy and Interpretability

Francisco Saraiva Pimenta Verdelhos

Master's Dissertation

Advisor at FEUP: Prof. Pedro Sanches Amorim
Advisor at the company: Mr. Brian Sieben

U. PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Master's in Industrial Engineering and Management

2024-06-27

Abstract

Supply chain (SC) design, planning, and operation decisions are critical to the success or failure of a company (Craighead et al., 2007). Measuring SC efficiency enables organizations to assess their operational performance, uncover enhancement opportunities, and refine their processes for improved outcomes. Despite the various existing approaches SC efficiency evaluation, no single method is universally recognized or suitable for all businesses. Companies have traditionally relied on methods such as Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA) to calculate SC efficiency (Cooper et al., 2003; Daraio and Simar, 2007; Burger, 2008). However, the limitations of both methods highlight the need for alternative approaches that can address these challenges effectively (Katharakis et al., 2014; Bauer et al., 1998). To bridge this gap, a research initiative by Sieben (2023) proposes a two-step method to quantify SC efficiency. Firstly, value curves are used to extract a proxy value for SC efficiency, which was named by Sieben (2023) as flow efficiency. Secondly, the derived flow efficiency is compared with the predicted flow efficiency.

This master thesis proposes a pioneering approach to improve flow efficiency prediction accuracy using machine learning (ML) models. Moreover, it provides visibility into the features that most significantly contribute to flow efficiency prediction, laying the groundwork for targeted improvements in underperforming SC.

To achieve the desired results, the methodology proposed in this master's dissertation begins with data consolidation, focussing on the extraction and preparation of both the target variable (theoretical flow efficiency calculated from value curves) and the explanatory variables (identified flow efficiency drivers). Following this, two primary approaches are employed. First, ML models are developed using Stepwise Backward Elimination (SBE) to explore the balance between model complexity and accuracy, aiming to identify flow efficiency drivers that are deemed the most important. Second, density-based spatial clustering of noise applications (DBSCAN) is utilised to identify distinct clusters of materials. These clusters are then analysed using various ML models to determine the best performing model cluster combinations. Finally, the importance of each feature within these combinations is assessed using SHAP values, providing insights into the factors influencing SC flow efficiency prediction.

The application of the described methodology to a case study developed within Hilti revealed room for significant advancements in supply chain management practices. The study identifies machine time and material cost as consistently critical features for predicting flow efficiency in various models. Moreover, clustering the data into subsets is proven not to significantly improve prediction performance, but to greatly enhance model interpretability. Clustering uncovers distinct patterns and highlights the variability in the impact of efficiency drivers across different clusters. This variability underscores the importance of considering cluster-specific characteristics for targeted SC efficiency improvements.

Resumo

As decisões de conceção e planeamento da cadeia de abastecimento (CA) são fundamentais para o sucesso de uma empresa (Craighead et al., 2007). A medição da eficiência da CA permite às organizações avaliar o seu desempenho operacional, identificar oportunidades de melhoria e aperfeiçoar os seus processos. Apesar das várias abordagens existentes à avaliação da eficiência nas CA, não existed ainda um método universalmente reconhecido. Tradicionalmente, as empresas têm-se apoiado em métodos como a Análise de Fronteira Estocástica (SFA) ou a Análise Envoltória de Dados (DEA) para calcular a eficiência das CA (Cooper et al., 2003; Daraio and Simar, 2007; Burger, 2008). No entanto, as limitações de ambos os métodos sublinham a necessidade de abordagens alternativas (Katharakis et al., 2014; Bauer et al., 1998). Para colmatar esta lacuna, uma iniciativa de investigação de Sieben (2023) propõe um método em duas etapas para quantificar a eficiência das CA. Em primeiro lugar, as curvas de valor são utilizadas para extrair um valor aproximado da eficiência da CA, que foi designado por Sieben (2023) como *flow efficiency*. Em seguida, a *flow efficiency* obtida é comparada com uma *flow efficiency* estimada.

Esta dissertação propõe uma nova abordagem para a melhoria da precisão da previsão da *flow efficiency* utilizando modelos de *Machine Learning* (ML). Além disso, pretende ainda dar visibilidade das características que contribuem de forma mais significativa para a previsão da *flow efficiency*, construindo uma base para uma melhor alocação de recursos na procura por uma CA otimizada.

Para alcançar os resultados pretendidos, a metodologia proposta começa com a consolidação dos dados, centrando-se na extração e preparação da variável-alvo (*flow efficiency* calculada a partir das curvas de valor) e das variáveis explicativas (*flow efficiency drivers* identificados). De seguida, são utilizadas duas abordagens distintas. Em primeiro lugar, são desenvolvidos modelos de ML utilizando *Stepwise Backward Elimination* (SBE) para explorar o equilíbrio entre a complexidade e a precisão dos modelo, com o objetivo de identificar os *flow efficiency drivers* considerados mais importantes. Em segundo lugar, utiliza-se *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) para identificar grupos distintos de materiais. Estes grupos são analisados utilizando vários modelos ML para determinar as combinações modelo-cluster com melhor desempenho. Finalmente, a importância de cada caraterística dentro destas combinações é avaliada utilizando valores SHAP.

A aplicação da metodologia descrita a um caso prático desenvolvido na Hilti revelou a existência de espaço para avanços significativos na gestão da cadeia de abastecimento. O estudo identifica o tempo de máquina e o custo do material como características críticas para a previsão da *flow efficiency*. Além disso, ficou provado que o agrupamento dos dados em subconjuntos não melhora significativamente a capacidade preditiva dos modelos, mas aumenta a interpretabilidade dos mesmos. O agrupamento revela padrões distintos e destaca a variabilidade no impacto dos factores de eficiência em diferentes contextos. Esta variabilidade sublinha a importância de ter em conta as características específicas dos clusters para melhorar a eficiência da CA.

Acknowledgements

To Mr. Brian Sieben, Head of Sourcing Excellence at Hilti, for your unwavering support and availability from the very beginning of this work. Thank you for challenging me, for letting me learn from my mistakes and above all, for teaching me the immense value of applying scientific principles in a business context. I would also like to extend my gratitude to the entire Sourcing Excellence department for providing me with room for personal and professional growth.

To Professor Pedro Amorim, my advisor at Faculdade de Engenharia da Universidade do Porto, for his guidance and constant motivation, encouraging me to push beyond my limits at every stage of this journey.

To Tiago Rodrigues, my dear friend and co-worker, for your invaluable guidance and knowledge shared not only over the past six months but also throughout the past few years.

To all my friends, for shaping me into the person I am today. A special thanks to the ones with who I had the pleasure of sharing unforgettable moments during these 5 years.

To my entire family for their unwavering confidence in me, for continually showing their pride in my achievements, and above all, for providing the emotional stability I need to fulfil my goals.

To my sister, for your support and for pushing me to be the best version of myself.

To my father, for teaching me the value of work, the wisdom in underpromising and overdelivering, and the importance of remaining true to our values even in times of great uncertainty.

To my mother, for your endless patience in listening to and supporting me, for understanding me better than I understand myself, and for believing in my potential more than I do. Thank you for always being where I need you the most, I hope I never let you down.

To my grandfather, for being my greatest inspiration. Thank you for your wisdom, pride, and care. I apologize that my selfishness is keeping me away from home. While I may not have become a doctor as you wished, I promise to always do my utmost to ensure the well-being of everyone around me.

Agradecimentos

Ao Sr. Brian Sieben, Diretor de Sourcing Excellence da Hilti, pelo seu apoio e disponibilidade desde o início deste trabalho. Obrigado por me ter desafiado, por me ter deixado aprender com os meus erros e, acima de tudo, por me ter ensinado o imenso valor da aplicação de princípios científicos num contexto empresarial. Gostaria ainda de estender o meu agradecimento a todo o departamento de Sourcing Excellence por me ter proporcionado espaço de crescimento pessoal e profissional.

Ao Professor Pedro Amorim, o meu orientador na Faculdade de Engenharia da Universidade do Porto, pela sua orientação e motivação constante, incentivando-me a ultrapassar os meus limites em cada etapa deste percurso.

Ao Tiago Rodrigues, amigo e colega de trabalho, pela sua inestimável ajuda e sabedoria partilhada não só ao longo dos últimos seis meses, mas também ao longo dos últimos anos.

A todos os meus amigos, por me terem moldado na pessoa que sou hoje. Um agradecimento especial àqueles com quem tive o prazer de partilhar momentos inesquecíveis nestes ultimos 5 anos.

A toda a minha família, pela confiança inabalável que sempre depositaram em mim, por demonstrarem continuamente o vosso orgulho nas minhas conquistas e, acima de tudo, por me proporcionarem a estabilidade emocional de que necessito para atingir os meus objetivos.

À minha irmã, por me motivar a ser a melhor versão de mim próprio.

Ao meu pai, por me ter sempre transmitido o valor do trabalho. Obrigado por me mostrares a importância de nos entregarmos a 100% aos objetivos que nos propomos a cumprir, e, acima de tudo, o benfício de nos mantermos fieis aos nossos valores independetemente do contexto que nos rodeia.

À minha mãe, pela inesgotável paciência em ouvir os meus desabafos, por saberes o que penso sem eu proferir uma única palavra e por muitas vezes acreditares no meu potencial mais do que eu próprio. Obrigado por estares sempre onde preciso de ti, espero nunca te desiludir.

Ao meu avô, por ser a minha maior inspiração. Obrigada pela tua sabedoria, orgulho e carinho. Peço desculpa por o meu egoísmo me manter longe de casa. Embora não me tenha tornado médico como desejavas, prometo fazer sempre o meu melhor para garantir o bem-estar de todos os que nos rodeiam.

"To develop a complete mind: Study the science of art; Study the art of science. Learn how to see. Realize that everything connects to everything else."

Leonardo da Vinci

Contents

Acronyms and Symbols

BU	Business Unit(s)
DEA	Data Envelopment Analysis
DMUs	Decision-Making Units DOS
Sourcing Excellence Department	
EDA	Exploratory Data Analysis
ERP	Enterprise Resource Planning
ML	Machine Learning
PMS	Performance Measurement System
SAP	System Analysis Program Development (Systemanalyse Programmentwicklung)
SBE	Stepwise Backward Elimination
SC	Supply Chain(s)
SCM	Supply Chain Management
SCPMS	Supply Chain Performance Measurement System
SFA	Stochastic Frontier Analysis
VBA	Visual Basic for Applications

List of Figures

List of Tables

Chapter 1

Introduction

This chapter serves as an introductory overview of the dissertation. It provides information about the context in which the motivation for developing this work emerged. The project's background will be presented by referencing previous relevant projects that will support this dissertation. Additionally, the research methodology and research questions will be detailed. To conclude, the structure of the dissertation will be outlined.

1.1 Context and Problem Scope

Supply chain (SC) design, planning, and operation decisions play a pivotal role in the success or failure of a company (Craighead et al., 2007). Measuring SC efficiency enables organizations to assess their operational performance, identify opportunities for enhancement, and refine their processes for improved outcomes. Although multiple approaches exist for evaluating SC efficiency, there is no single method that is universally recognized or suitable for every business.

Companies have relied on methods such as Stochastic Frontier Analysis (SFA) - a parametric approach, which assumes a specific functional form and statistical distribution for the data - or Data Envelopment Analysis (DEA) - a non-parametric approach - to calculate and identify opportunities for efficiency improvements in their operations (Cooper et al., 2003; Daraio and Simar, 2007; Burger, 2008). However, SFA requires normally distributed data while DEA struggles to handle noise in the data. (Katharakis et al., 2014; Bauer et al., 1998). Therefore, there is room for alternative approaches that can address these challenges.

A recent pioneering research initiative proposing a novel approach to quantify efficiency has been conducted by Sieben (2023) to bridge the identified gap. Initially, value curves are plotted to depict the relationship between value and time at each stage of production. These curves facilitate the computation of a proxy for SC efficiency, which was named by Sieben (2023) as flow efficiency. A value curve example and the base concept for efficiency calculation are presented in 1.1 In the second step, the derived flow efficiency is compared against a flow efficiency predicted through ML models.

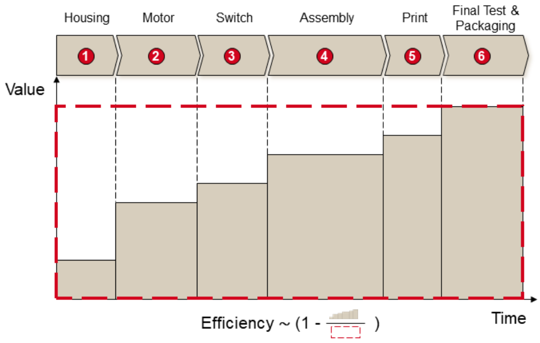


Figure 1.1: Value curve example and efficiency calculation (Parlak and Pescalli, 2023).

This master's dissertation aims not only to enhance the methodologies for flow efficiency prediction but also to provide visibility on the parameters that are impacting flow efficiency the most. A primary emphasis will be placed on data extraction and processing, as this is crucial for the precise identification of drivers affecting flow efficiency. Furthermore, the study will focus on the development and evaluation of ML models designed to quantify SC efficiencies.

The anticipated outcomes of this research should offer insightful visibility into materials that are underperforming. As a result, it will facilitate targeted interventions for the enhancement of SC operations, ensuring a strategic alignment between the SC configurations and both market and operational demands.

The motivation behind this project is to validate the model proposed by Sieben (2023) as a reliable alternative to traditional methodologies for calculating SC efficiency. To test the applicability of this methodology, Hilti's business context is used as a case study, leveraging its data to address the issue of SC strategic alignment. This approach not only aims to contribute to the scientific community but also demonstrates practical applications in a real-world business environment.

1.2 Research Method and Heuristic Framework

This dissertation will follow the research method proposed by Ulrich (1981), as described in the Dissertation Structure section.

This masters' thesis is going to leverage the research project conducted by Sieben (2023) which encompasses bachelors' (Näf (2015); Wolf (2020)) and master's theses (Parlak and Pescalli (2023); Gomes (2021); Klein (2021); Putkivaara (2020); Musacchia (2019); Thampi (2018); Rodrigues (2023)), a journal paper (Sieben et al. (2023)) and a doctoral thesis (Sieben (2023)). Rodrigues (2023) findings on the best-performing flow efficiency calculation methods are going to be directly used in this project. Moreover, Parlak and Pescalli (2023) contributions on the impact of ML on flow efficiency prediction will be considered during this dissertation's methodological development.

Hilti employs performance pricing (VDI, 2015) within its sourcing departments as a well-established method and is now seeking enhancements on flow efficiency calculations by applying similar practices. Given that flow efficiency remains an emerging area of research, ML techniques stand out as promising approaches as they are expected to enable Hilti to derive actionable insights from their flow efficiency-related data, facilitating data-driven decision-making. Thus, Hilti's interest in improving internal processes allowed us to investigate flow efficiency calculations.

The methodologies proposed throughout this dissertation will be applied to datasets extracted from Hilti's databases. This data will be used to evaluate the reliability and assess the performance of the described methods. After identifying the model or models that better suit the context under analysis, the outcomes of the methodological development will be implemented in a specific case study provided by Hilti.

1.3 Research Questions

A research question is designed to precisely delineate the issue that will be the focal point throughout a research endeavour. In the case under analysis, the main focus will be the improvement of flow efficiency prediction.

The accuracy of predicted flow efficiency is highly dependent on the quality of the input data. To address this challenge, artificial intelligence, in particular ML techniques, is used to support pattern identification and provide insights into how different drivers affect SC flow efficiency in various contexts. After an improved selection of drivers, it is possible to use ML models to increase the accuracy of flow efficiency predictions and extract actionable insights from these results.

The accuracy of predicted flow efficiency heavily relies on the quality of input data, specifically the data on flow efficiency drivers. Establishing a strong foundation with reliable data enables the use of ML techniques to identify patterns and understand the impact of various drivers on flow efficiency predictions. By exploring alternative approaches, it is possible to enhance prediction accuracy while ensuring model interpretability, thus allowing for the extraction of actionable insights from the results.

Given the context outlined and the benefits of not only refining existing models for predicting flow efficiency but also evaluating their comparative performance, the research questions are:

- How can flow efficiency drivers be identified for a more accurate flow efficiency prediction?
- How can machine learning be leveraged to improve the accuracy of SC flow efficiency prediction?
- How do different flow efficiency drivers affect flow efficiency prediction?

1.4 Dissertation Structure

The structure of the dissertation follows the methodology proposed by Ulrich (1981), chosen for its effectiveness in addressing real-world issues in a rigorous scientific way (Ulrich, 1982). Chapter 1 introduces the subject of this thesis by presenting the heuristic framework, the problem statement, the research method and the research questions.

Considering the research questions introduced in Chapter 1, Chapter 2 focuses on a comprehensive review of the literature on SC efficiency drivers and SC efficiency calculation techniques. This review clarifies the current state of the art and pinpoints existing gaps in the field. Chapter 3 presents the core idea of this thesis, building upon the gaps identified in Chapter 2 and outlining the formal methods to be explored for addressing the research questions.

Chapter 4 is dedicated to an in-depth exploration of the methodology followed to extract data and systematically identify drivers of Flow Efficiency.

Building on the findings from Chapter 4, Chapters 5 and 6 explore two distinct approaches for predicting flow efficiency and evaluating feature importance. Chapter 5 assesses various ML techniques using different sets of flow efficiency drivers. The objective is to identify a set of drivers that consistently yield accurate predictions. Conversely, Chapter 6 concentrates on evaluating the importance of different flow efficiency drivers in the prediction process in subsets of the original data.

Moreover, in Chapter 7, the described methodology is applied to a specific case study to validate the reliability of the work developed. Hilit's company profile is shared to help the reader better understand the context in which this thesis was developed.

Chapter 8, Discussion and Outlook, presents the answers to the research questions and this dissertation's contribution to science and business. It also concludes by discussing the limitations faced and the avenues for further research.

The adaptation of Ulrich's research structure to this work is shown in Figure 1.2:

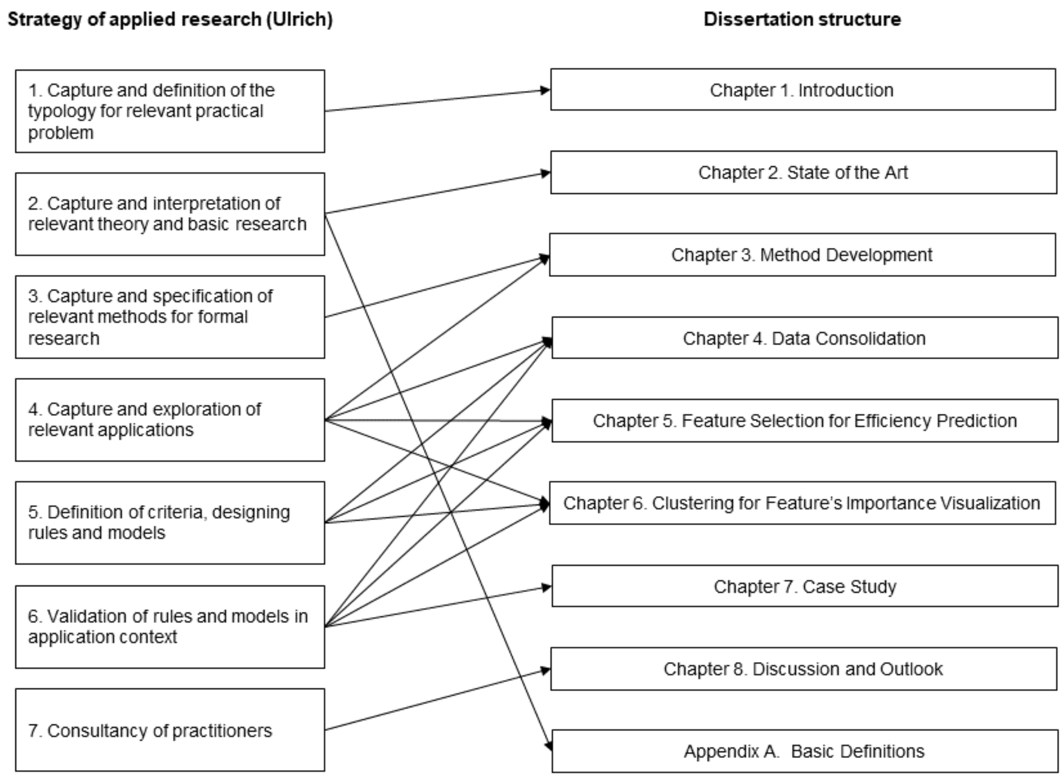


Figure 1.2: Dissertation's structure (Ulrich, 1981) (modified).

Chapter 2

State of the Art

This chapter builds on the foundational concepts from AppendixA and the research questions outlined in Chapter 1 to thoroughly analyse the current research landscape. The conducted research is divided into three primary focus areas. Initially, the study explored the drivers of efficiency in SC. Subsequently, an examination of performance measurement systems (PMS) within the context of dynamic SC is undertaken. Finally, research on the evaluation methods for the variable importance in the context of ML models is conducted.

2.1 Efficiency Drivers for Flow Efficiency Prediction

Efficiency drivers play a crucial role in shaping the overall efficiency of a SC. These drivers encompass a range of factors that impact various aspects of SC's operations. Businesses can optimize their SC processes and enhance their competitiveness, by understanding and effectively managing these drivers.

This section is split between two main areas. First, a comprehensive literature research on the factors that impact SC efficiency is conducted. Secondly, the author focused on finding techniques that allow the selection of the most important variables among a predefined range of parameters, tackling high dimensionality challenges.

2.1.1 Supply Chain Efficiency Drivers Identification

This section predominantly draws from the research undertaken by Niklas Bley during his Master's dissertation (Bley, 2022). Bley (2022) aimed to pinpoint the factors influencing SC efficiency through a comprehensive literature review complemented by parameters deemed as critical by Hilti for evaluating SC efficiency. To ensure a comprehensive representation of all aspects of the SC the SCOR model is applied, ensuring each facet of the SC is accounted for by at least one efficiency driver, thus offering a holistic perspective.

The literature review is a two-step process. Firstly, Bley (2022) specifically searched for articles that examine the key metrics used by Hilti. Secondly, a conventional literature search is conducted focusing on SC and SCM. From this body of literature, additional sources are identified that either corroborated statements from other works or introduced new efficiency drivers.

The identified SC efficiency drivers are presented in 2.1

Table 2.1: Supply chain flow efficiency drivers (Bley, 2022).

Driver	Driver
Annual demand	Product commonality
Batch size	Product complexity
Business interruption risk (BI)	Product value density
Complexity of supplier's network	Purchase price
Customer order decoupling point (CODP)	Response time
Delivery reliability	Selling price
Flexibility of supplier's network	Throughput
Forecast error	Throughput variance
Inventory	Transportation cost
Lead time	Turnover
Lead time variance	Unit production cost
Minimum order quantity (MOQ)	Weight and volume
Obsolescence risk	Working capital

To validate the findings of Bley (2022), a new Scopus search is conducted. Initially, the query "Supply chain efficiency drivers" yielded no results. Assuming the lack of literature is due to specific terminology, the query is revised to "Supply chain performance factors," which resulted in finding 8 relevant articles. These articles delved into comprehensive literature reviews on the various factors directly or indirectly impacting SC efficiency. The majority of the previously identified flow efficiency drivers are corroborated through this validation process. However, considering the growing significance of technology in contemporary SC processes, it is deemed essential to mention other authors and incorporate Technological Integration as a new efficiency driver.

Technological integration refers to the strategic utilization of digital technologies to connect, automate, and optimize information flow and collaboration throughout the entire SC. Studies have shown that both digitalization and SC integration positively influence firm performance (Liu and Chiu, 2021).

Additionally, to better illustrate the dynamic nature of supply-demand relationships across different product maturity stages, the concept of Product lifecycle is also important to pinpoint.

Product lifecycle encompasses the phases a product traverses from its introduction to the market until its eventual withdrawal or discontinuation. These stages typically include introduction, growth, maturity, and decline. SC strategies must dynamically adapt to ensure alignment with product characteristics and customer requirements, thereby maximizing competitiveness (Aitken et al., 2003).

This review process tries to ensure the preservation of the holistic and contemporary nature of the parameters under analysis.

2.1.2 Supply Chain Efficiency Drivers Selection

In this section, techniques are explored to mitigate the challenges posed by high-dimensional datasets on predictive models for SC efficiency. Given the narrow scope of available research, Scopus yielded no results. Consequently, the works of Bley (2022) and Parlak and Pescalli (2023) serve as state-of-the-art references for feature selection impacting SC efficiency prediction.

Bley (2022) emphasizes the criticality of ensuring data availability as a preliminary step in selecting flow efficiency drivers for analysis. To streamline resource allocation, an individual evaluation of each driver based on predefined criteria is performed. The high-level criteria include the presence of data within the company and the necessity of data collection. As subcategories, Bley (2022) uses the VDI 2817 standards referenced for assessing data quality VDI (2015). If data is deemed missing or unavailable, a cost-benefit analysis is suggested, considering factors such as the effort and cost of data collection alongside the relevancy of the efficiency driver.

Following this assessment, a comparison of results across different products or product families enables the selection of drivers for analysis, ensuring robust feature selection for effective SC efficiency prediction.

Building on this methodology, Parlak and Pescalli (2023) focuses on the usage of EDA and ML techniques to identify the most important variables for more accurate analysis in flow efficiency calculations. The authors take into account the previously identified pool of drivers and apply two different methodologies. On the one hand, correlation matrices are used to gain visibility on the relationships and patterns of the explanatory variables. This technique enabled a selection of the parameters that mostly impact the prediction of the SC flow efficiency. On the other hand, PCA is used to further reduce the dimension of the dataset that would later be fed to the ML models.

2.2 Methods for Production and Logistics Performance Estimation

The research presented addresses the necessity of enhancing SC performance to meet customer expectations. The central focus of achieving this improvement is the adoption of effective performance measures and metrics that assess the efficiency of the SC. An effective PMS is therefore critical to the success of any business, enabling the measurement of relevant indicators at opportune times.

Over the past few decades, numerous researchers have conducted literature reviews on supply chain performance measurement systems (SCPMS). However, most of these studies have integrated SC as components of broader PMS frameworks. In the current market environment, the adoption of SCM approaches and techniques is becoming increasingly prevalent (Vitasek, 2013). This underscores the need for a comprehensive review of PMS in the specific context of SC.

Thus, the main focus is the review of existing literature on PMS within the dynamic SC environment and the identification of potential research gaps that could be explored in this dissertation.

To fulfil this objective, the author has employed a systematic literature review procedure. Initially, the author utilised databases such as Scopus, Google Scholar and ISI in order to collect literature from the year 2000 to 2024. Over 900 papers are identified in the context of SCPMS, utilising the keywords "supply chain performance", "supply chain efficiency measurement", "supply chain performance measurement" and "supply chain performance methods". Initially, redundant papers collected from different databases are eliminated. To improve the quality of the papers under analysis, the title, keywords, abstract and conclusions are reviewed. Ultimately, 127 papers are considered to be included in the review process in the context of SCPMS.

After the papers' analysis, the following techniques are found to be the most relevant.

2.2.1 Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) is a comprehensive theory of measurement that derives ratio scales from both discrete and continuous paired comparisons. These comparisons can be based on actual measurements or on a fundamental scale that reflects the intensity of preferences and perceptions. AHP is particularly focused on measuring inconsistencies, as well as examining dependencies within and between the groups of elements in its structure (Saaty, 1987). This technique has been extensively applied in fields such as multi-criteria decision-making, planning, resource allocation, and conflict resolution.

Even though it was developed in the 1970s, only in the early 2000s the model is applied in the SCM context.

Chan (2003) leverages AHP to address the challenge of prioritizing performance measures in SCM. This methodology facilitates an objective evaluation of each performance measure by pairwise comparisons. This process is a key innovation in the paper, providing a clear and quantifiable way to balance several SC efficiency drivers. The author offers a detailed guide on how to methodically evaluate and rank the importance of different performance measures, thus enabling managers to make informed decisions about where to focus their improvement efforts.

Cho et al. (2012) presents a framework for measuring the performance of service SC, which is particularly groundbreaking given the limited exploration of this area in existing research. The innovation lies in integrating fuzzy logic with the traditional AHP, enabling a more effective handling of the ambiguities and subjective evaluations common in performance assessments. By setting a priority between the different dimensions, the framework not only offers a methodological advance but also provides a practical tool for SC managers to enhance efficiency.

2.2.2 Supply Chain Operations Reference

The SC Operations Reference (SCOR) model serves as a framework for evaluating and improving SC management practices (Stevens, 1997). This model categorizes all SC activities into six core management processes: Plan, Source, Make, Deliver, Return, and Enable. By integrating these elements, the SCOR model facilitates detailed mapping of SC operations, benchmarking against industry standards, and the identification of areas needing improvement.

The SCOR model's strength lies in its broad applicability across diverse industries, providing tools for businesses to achieve operational efficiency. It quantifies performance using standardized metrics based on reliability, responsiveness, agility, costs, and asset management efficiency, which guide strategic decision-making and operational adjustments (Stevens, 1997).

2.2.3 Stochastic Frontier Analysis

SFA is a parametric approach that aims to assess the efficiency of production units. This method incorporates both random errors due to external factors and inefficiencies related to the production process itself (Aigner, 1977). SFA operates by estimating a frontier production function which serves as the benchmark for maximum possible output given a set of inputs. The efficiency of each decision-making unit (DMUs) - any entity that is to be evaluated in terms of its abilities to convert inputs into outputs - is then evaluated based on its distance from this frontier, with consideration for stochastic variations that could skew the measurement of output inefficiencies. SFA is particularly beneficial in contexts where data might be influenced by noise and measurement errors. The methodology not only identifies the presence of inefficiencies but also quantifies them, thereby offering critical insights for performance improvement and strategic management across diverse contexts.

Hamdan et al. (2017) presents a methodology based on SFA to analyze the efficiency of the Build-to-Order SC (BTO-SC), contrasting it with that of traditional SC. The findings aim to provide a robust analytical framework that can serve as an investment guideline for companies considering the adoption of BTO-SC principles, offering insights into the potential benefits and efficiencies gained from this agile SC model.

2.2.4 Data Envelopment Analysis

DEA is a non-parametric method for estimating production frontiers and assessing the relative efficiency of DMUs without assuming a specific functional form of the production process (A Charnes and Seiford, 1997). The original DEA model evaluates units based on their inputs and outputs using mathematical programming to form an efficiency frontier. This approach compares each unit against a combination of others to determine an efficiency score, indicating whether units are efficient (on the frontier) or inefficient (below it). DEA's strength lies in its ability to handle multiple input and output scenarios without a predefined production function, making it versatile for performance evaluation across various fields. The DEA model has been extended to address specific conditions and incorporate additional variables, establishing it as a crucial methodology in operations research and performance management. Literature research indicates that DEA is commonly used in SC performance measurement.

Wong and Wong (2007) examines the application of DEA to assess internal SC performance, with a particular focus on two distinct models: the technical efficiency model and the cost efficiency model. The study highlights how DEA helps identify inefficiencies within SC operations and guides managers in implementing effective remedial actions and resource allocation strategies.

Tavana et al. (2013) proposes an extension of DEA models using Epsilon-Based Measures (EBMs) of efficiency, which allow for the simultaneous consideration of both radial - scale by the same factor - and non-radial inputs and outputs. The new model termed the Network EBM (NEBM), integrates these measures into a unified framework to address network DEA problems effectively.

2.2.5 Machine Learning

Considering that ML will be used to address the research questions posed in Chapter 1, the author decided to narrow the scope of the research. To comprehensively investigate the usage of ML algorithms to estimate the efficiency of SC a new Scopus research is conducted. Using "machine learning", "supply chain", and "performance measurement" as keywords, 7 papers are found. Taking into account the low amount of articles available, every single one is analysed. Table 2.2 summarizes the reviewed articles.

The scarcity of documentation on this specific aspect underscores a clear gap in the literature. This signals an opportunity for further investigation, emphasizing the unexplored potential of ML methodologies in predicting SC efficiency

2.3 Review of literature

This section provides insights into existing studies related to the subject of this dissertation. The author has carefully selected papers using keywords relevant to the method development discussed in the following chapters. The work is categorized into review areas, which define the problem-relevant context, and design areas, where systematic changes are proposed. In the Scope of the Review and Design Table (Figure 2.1), keywords are listed in the first row and marked with hollow or solid circles to indicate whether they are partly or fully covered.

		Areas of Review				Areas of Design		
		Efficiency and Productivity Measurements	Value Curves	Flow Efficiency Calculation	Literature Based Efficiency Drivers Identification	Flow Efficiency Drivers Identification with ML	Flow Efficiency Prediction with ML	Feature Importance for Flow Efficiency Prediction
Nienhaus	2004	•	•					
Näf	2015	•	•	•				
VDI 2817	2017			•	•			
Thampi	2018	•	•	•				
Rehman et al.	2020	•			◦			
Klein	2021	•	•					
Gomes	2021	•	•	•	•			
Bley	2022	•	•	•	•			
Parlak & Pescalli	2022	•	•	•	•	◦	•	
Rodrigues	2023	•	•	•				
Sieben	2023	•	•	•	•			
This Work	2024	•	•	•	•	•	•	•

Figure 2.1: Comparison of the areas addressed by former researches with the ones targeted by this work.

• Fully Addressed
◦ Partially Addressed

Table 2.2: Machine learning on supply chain performance measurement.

Author	Model Used	Context	Takeaways
Nilashi et al. (2024)	K-means clustering and fuzzy logic	Evaluating the performance of Electric Vehicles SC by analyzing indicators related to customer perceived value.	ML-based method can accurately predict Electric Vehicle SC efficiency.
Farchi et al. (2023)	Multidimensional performance measurement model and Artificial Neural Networks (ANNs)	Quantifying sustainable performance in the road freight transport sector across economic, social, environmental, operational, and stakeholder dimensions.	The model is general and applicable to various disciplines, leveraging ANNs to predict global performance.
Dixit and Gupta (2023)	ML (general application)	Identifying critical parameters for quality control and predicting SC performance.	ML helps improve process efficiency by focusing on critical factors, though detailed methodologies are not discussed.
Mariappan et al. (2023)	Ensemble of regressors and classifiers	Predicting processing and shipment times for medical supply orders in an e-Pharmacy post-COVID-lockdown scenario.	ML techniques significantly improve prediction accuracy for order processing and shipment times.
Kliangkhlao et al. (2022)	Causal Bayesian Networks (CBNs)	Understanding market dynamics, balancing demand and supply, and assisting decision-makers in managing the SC.	CBNs provide a human-like approach to explaining demand and supply events, aiding in decision-making.
Ghaouta et al. (2021))	ML (general application)	Predicting warehousing efficiency by modeling warehouse operations and KPIs.	ML can extract generic knowledge of warehouse operations, enabling future scenario predictions.
Shirota et al. (2021)	Shapley values	Relating operational competencies to stock price evaluation by using operational measures as explanatory variables.	Using Shapley values provides more reliable information about management quality competence.

Chapter 3

Method Development

This chapter aims to describe the concept behind the developed methodology as well as the formal methods that are used to answer the research questions posed in Chapter 1. This chapter is essential for the understanding of the subsequent ones.

3.1 Approach to Concept Development

Chapter 2.1.2 highlights ongoing studies on flow efficiency prediction and the application of performance pricing to derive actionable insights from flow efficiency values. Parlak and Pescalli (2023) contributed significantly to this field by using ML to identify flow efficiency drivers and improve prediction model performance. Despite their contributions, inconsistencies in the selection of flow efficiency drivers across various contexts and significant data collection challenges suggest that there is still room for improvement. A promising future research direction identified by Parlak and Pescalli (2023) involves applying cross-validation across different product families. This approach consists of training ML algorithms with one product family and testing them with another to ensure robust validation. Achieving this requires large datasets with consistent explanatory variables and sufficient sample sizes.

Building on their work, this dissertation aims to study feature importance, hypothesizing that the predicted flow efficiency is influenced by different drivers for different materials. Additionally, it seeks to leverage ML capabilities to enhance the accuracy of flow efficiency predictions. The focus will be on developing a methodology that is both generalizable across diverse contexts and ensures the interpretability of results.

The diagram presented in 3.1 describes the several stages executed during the methodological development to answer the research questions posed in Chapter 1.

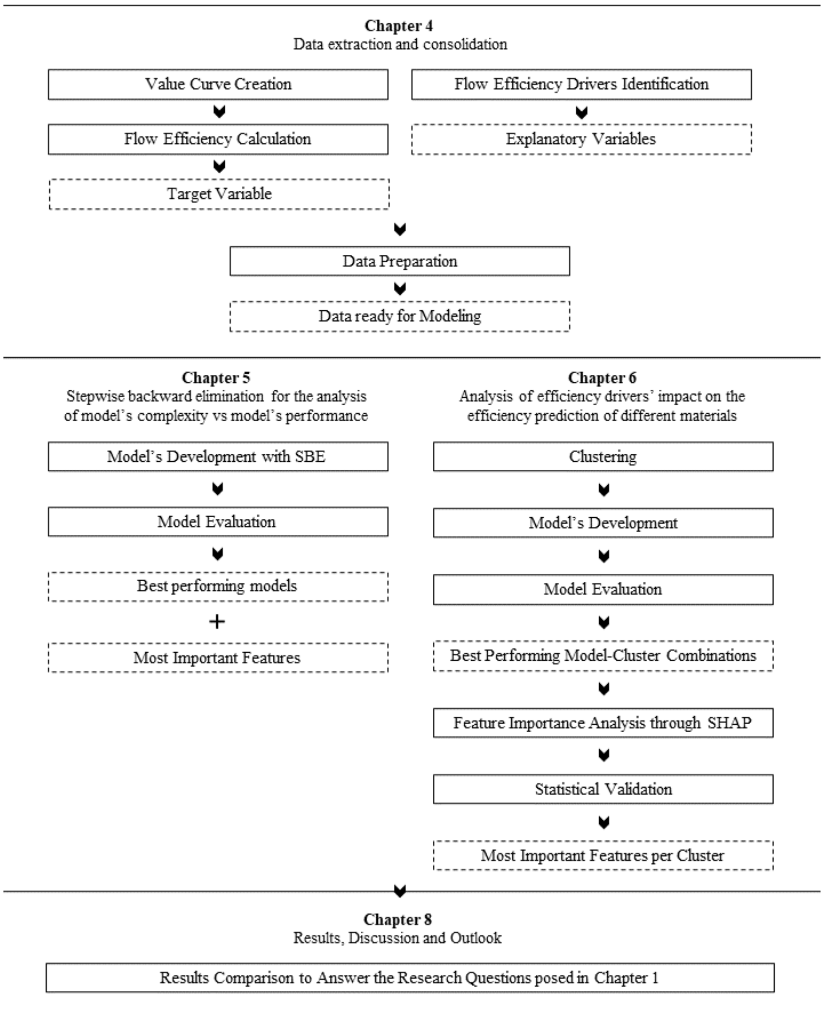


Figure 3.1: Concept development stages (The author).

The methodology begins by defining the materials within the scope of analysis for which value curves will be created. Previous master dissertations conducted in this or similar contexts within Hilti relied on manual tasks to extract the data required for flow efficiency calculation through the value curves, limiting the amount of data that could be gathered and potentially compromising ML model performance. To address this, a data extraction automation is implemented to mitigate the impact of data scarcity on model performance. The value curves are then used to calculate the theoretical flow efficiency used as the target variable for ML models.

In parallel, a range of flow efficiency drivers are also identified and extracted from Hilti's Enterprise Resource Planning (ERP) to be fed as explanatory variables to the ML models. After extracting all the required data, it goes through several steps of preparation to ensure improved model performance.

From this point onwards two different approaches are followed with the objective of testing different hypotheses. In Chapter 4, ML models are developed alongside Stepwise Backward Elimination (SBE). The entire dataset is used to explore the balance between model complexity and accuracy. This investigation aims to determine if there is a specific set of features that consistently emerge as the most important for predicting flow efficiency.

In Chapter 5, clustering is performed to build a ground that allows to testing of the hypothesis that different materials have distinct drivers influencing the flow efficiency of their SC. Clustering the data comes alongside the objective of providing visibility on the most relevant parameters in different subsets of the original data. Subsequently, different ML models are applied, considering the specificities of the dataset under analysis. These are evaluated based on predefined criteria to identify the best-performing model-cluster combination. To conclude, the importance of each parameter for the predictive capacity of the models is assessed to test the hypothesis that the flow efficiency drivers impacting SC efficiency vary across different materials.

The results from both approaches are then evaluated and compared in Chapter 8 in order to provide clear answers to the research questions posed in Chapter 1. This multifaceted analysis not only enhances the predictive performance but also offers valuable insights into the specific factors affecting SC efficiency. It provides diverse perspectives on the drivers impacting these predictions and how they influence the outcomes. By leveraging both feature selection techniques and clustering, the study systematically examines model complexity, accuracy and interpretability.

3.2 Formal Methods

In this subchapter, formal methods essential for methodological development are introduced. These methods are crucial to ensure that the resulting model yields accurate results, enabling precise answers to the research questions.

3.2.1 Data Consolidation

This subsection outlines the initial steps of the methodological development aimed at consolidating the necessary data to predict SC flow efficiency. Firstly, Visual Basic for Applications (VBA) is employed to optimize the extraction of data required for the value curve creation. After drafting the value curves, the theoretical flow efficiency is calculated.

Next, the focus shifts to identifying and select flow efficiency drivers, a crucial aspect of methodological development, given their role as explanatory variables in predicting flow efficiency. To ensure precise identification, the author builds upon the methodology outlined by Bley (2022).

Finally, during the data preparation stage, data quality is validated using the criteria described in VDI (2015). Based on the results of this evaluation, the extracted data undergoes cleaning, transformation, and consolidation. EDA techniques, such as correlation matrices, are employed to reduce multicollinearity in the dataset, thereby decreasing its dimensionality.

3.2.2 Feature Selection for Supply Chain Flow Efficiency Prediction

The methodology for predicting SC efficiency encompasses several stages, each leveraging ML capabilities to enhance prediction accuracy and develop a generalizable approach.

In the prediction stage, ML models are carefully chosen to align with the data characteristics. Two key considerations guided the selection of the most suitable ML models.

Firstly, acknowledging the high degree of multicollinearity in the dataset, Partial Least Squares (PLS) regression is selected. Multicollinearity refers to the presence of high correlation among the independent variables, which can lead to unstable and unreliable estimates of the regression coefficients in traditional multiple regression analysis (Chatterjee and Hadi, 2015). PLS regression is able to handle data with high collinearity, noise, and numerous independent variables, improving predictive performance by allowing all efficiency drivers to influence the dependent variable while considering inter-variable influences (Vinzi et al., 2010).

Secondly, recognizing that the target variable, flow efficiency, is a form of bounded data — since it is calculated through value curves and inherently limited to a range between 0 and 1 — Beta regression is chosen. This regression is particularly well-suited when the response component is restricted to an interval (0, 1), such as proportions, percentages, and fractions (Ferrari and Cribari-Neto, 2004).

To complement the pool of models identified based on the specific characteristics of the dataset, commonly used ML models are also incorporated into the methodological development.

SBE is used while developing the identified ML models. This method aims to evaluate the models' performance with different sets of features and investigate the possibility of reducing model complexity. It is essential to understand if there is a specific set of features consistently yielding accurate results for predicting flow efficiency. Through the feature importance calculation embedded in the SBE approach, it is possible to draw conclusions on this matter. This method is specifically chosen because it helps reduce multicollinearity by evaluating each predictor in the context of the others, thereby minimizing the presence of highly correlated variables in the final model. In contrast, Stepwise Forward Elimination (SFE) adds variables one at a time based on their individual predictive power, which can inadvertently include highly correlated predictors early in the process. Taking into account the likelihood of the dataset under analysis having a high degree of multicollinearity SBE is used, ensuring that the final model is less prone to issues arising from correlated features(Theng and Bhoyar, 2024).

For each one of the identified models, and for each different set of features, K-fold cross-validation is employed to mitigate overfitting risks, enhancing the overall reliability of the results.

3.2.3 Clustering for Feature's Importance Visualization

As one of the main goals of this dissertation is to provide interpretability of the results, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is employed as a clustering technique in a separate approach. Unlike many clustering algorithms that require the number of clusters to be specified in advance, DBSCAN does not. This feature makes it highly suitable for exploratory data analysis where the number of clusters is not known beforehand (Schubert et al., 2017). DBSCAN identifies clusters based on the density of points, allowing it to effectively identify clusters of arbitrary shapes and handle noise and outliers (Schubert et al., 2017). By employing this technique, the study aims to test whether the flow efficiency of different materials is influenced by different flow efficiency drivers.

After clustering, the data is used to train the aforementioned ML models, aiming to find the best cluster-model combination. After evaluating and identifying the best-performing model for each cluster, the author focuses on testing the hypothesis that the importance of flow efficiency drivers varies in predicting the flow efficiency of materials across different clusters. By calculating SHAP values for each model within the different clusters, it is possible to quantify the contribution of each flow efficiency driver to the model's predictions. This approach enables a detailed comparison of feature importance across clusters, highlighting any variations in the influence of specific drivers on flow efficiency predictions.

To strengthen the statistical validity of these findings, the author first uses the Shapiro-Wilk test to evaluate if the data is normally distributed. Building on those findings, ANOVA tests and the Kruskal-Wallis test are used to evaluate the significance of the differences in the SHAP values between clusters. This rigorous approach ensures that any observed differences in feature importance are statistically significant, providing robust evidence to support the hypothesis.

Chapter 4

Data Consolidation

This chapter provides a detailed description of the data consolidation methodology introduced in Chapter 3. It involves the description of the value curve creation process, the systematic identification of flow efficiency drivers and data preparation techniques used to adjust the dataset with the purpose of increasing ML models' performance and reliability.

4.1 Value Curves Creation

Value curves are used to calculate SC flow efficiency, which then serves as the target variable for the ML models. Some value curves drafted in the context of previous dissertations conducted at Hilti were already available for analysis but were not enough to draw meaningful conclusions. Therefore, the author chooses to focus on automating the creation of value curves, aiming to mitigate the risk of data scarcity negatively impacting the model's predictive performance. In general, the performance of ML models is highly dependent on the quantity and quality of the data used for training. More data typically leads to better model accuracy and robustness, as it allows the model to learn more patterns in the data, thus improving its predictive capabilities (Halevy et al., 2009).

The described procedure is an improvement of a method developed by Ege and Zeno in their dissertation (Parlak and Pescalli, 2023). Through the ck13n transaction in *Systemanalyse Programmentwicklung* (SAP), Hilti's ERP, it is possible to retrieve a detailed report of the steps involved in manufacturing processes, including operations' time and cost.

The value curve extraction method is summarized in the following steps:

- **Step 1: Identify materials in scope for the analysis**

  The methodological development faces limitations due to the unavailability of data necessary for the creation of value curves. Since this process relies on operational time and cost data for each step, our analysis is confined to materials produced in-house, as information for externally sourced products is unavailable. During the initial scoping phase, a vast array of 21,829 distinct materials were flagged as available for potential inclusion in this analysis

- **Step 2: SAP data collection**

  A Consolidated Standard Cost (KSK) transaction is performed in order to retrieve a report containing both time and cost data. For each manufacturing operation extracted, two types of elements are present:

  - The name of the operation along with its cost and time.
  - The components/raw materials name, cost and quantities.

  Moreover, general cost components such as the material, production and administrative overhead are also retrieved.

  This step was the primary bottleneck in the data extraction process. For each material number, a SAP transaction is executed, followed by the manual download and analysis of an Excel file, making it an extremely time-consuming task. To address this challenge, a VBA script is developed. Given a list of material IDs, it is possible to output the required cost and time data in the appropriate format for the next step.

  From the initial array of materials, only 8,764 have the desired data available for extraction. Therefore, the other 13,065 are excluded from the analysis.

- **Step 3: Cost and time data calculation per operation**

  For each operation the time and cost are calculated as follows:

  - Time Calculation: The duration of each production step, which coincides with the operations, is determined by summing up the time of each component involved in the operation.
  - Value Calculation: The cost of each production step is calculated by summing the cost of the components involved in that step along with the cost of running the machine.

- **Step 4: Value curves drafting**

  The value curves are created based on production steps and their associated costs. Out of the 8,764 materials with extracted data, only 6,595 contained the necessary information for value curve creation, meaning that for the remaining 2,169 materials, only data on either production time or production cost is available. Since both parameters are required to draft value curves, these materials were excluded from the analysis.

The last two steps are executed through a VBA script developed by Rodrigues (2023) during his dissertation and further adjusted to accommodate products with up to 50 production steps.

4.1.1 Flow Efficiency Calculation

The flow efficiency calculation is determined using vertical slack-based weighting. In his dissertation, Rodrigues (2023) evaluated 25 different methods for calculating flow efficiency and statistically demonstrated that this method was the most consistent for the desired context. Consequently, this study builds on that work, utilizing the vertical slack-based weighting formula to compute flow efficiency values. These values serve as the target variable for the ML models.

4.2 Flow Efficiency Drivers Identification

Choosing the appropriate flow efficiency drivers to be included in the analysis is an indispensable step, as these parameters will serve as explanatory variables for the ML models, significantly impacting their overall performance.

4.2.1 Data Availability

The first step involves analyzing the outputs from the literature research conducted on this topic. Considering the list of flow efficiency drivers provided by Bley (2022) in his dissertation, along with additional parameters identified by the author, the initial pool comprises 29 different efficiency drivers.

It is crucial to assess which efficiency drivers can be incorporated into the modelling process, as companies often lack complete data for each one of the identified drivers (Wang et al., 2016). Determining which ones to include or omit before data collection ensures that consistent efficiency drivers are available as independent variables for all products under review. This approach aims to save time and costs by avoiding the collection of irrelevant data.

In order to identify which flow efficiency drivers should be included in the analysis the methodology described by Bley (2022) is followed. Bley (2022) suggests the creation of the matrix shown as an example in Figure 4.1. One axis of the matrix contains the efficiency drivers from chapter 3.2.1 while the other axis contains the criteria used to evaluate whether an efficiency driver should be included or not. Bley (2022) suggests the higher-level criteria to be whether data is already available in the company and/or whether data needs to be collected. These primary criteria are then be subdivided into further sub-criteria that refer to the parameters used for data quality assessment from VDI 2187 (VDI, 2015). If the assessment identifies that data is missing or generally not available, it is evaluated whether data collection makes sense through a cost-benefit evaluation.

To ensure a thorough and reliable assessment, leveraging the expertise within the company where this dissertation is conducted, several meetings were held with key stakeholders, including materials managers, product managers, and material master experts. These collaborative sessions were crucial for accurately completing the evaluation matrix while guaranteeing visibility into the available data. This approach not only enhances the reliability of the data assessment but also streamlines the identification of relevant efficiency drivers.

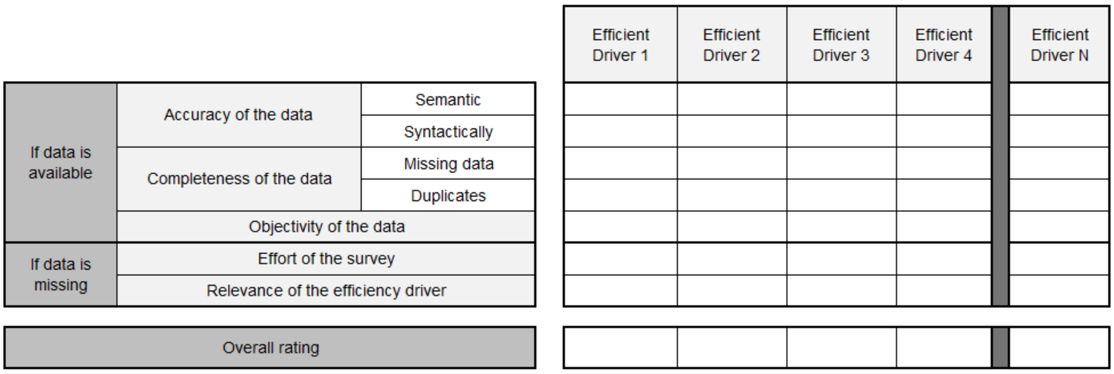


Figure 4.1: Example of the matrix for efficiency drivers selection(Bley, 2022).

Based on the insights gathered from these experts, a new evaluation criterion - Data Suitability - is added to the matrix. This criterion ensures that the efficiency drivers are relevant to the context under analysis, encompassing the fact that only in-house-produced products are in scope. Drivers deemed unsuitable are left out of the analysis.

The final evaluation matrix detailing the inclusion criteria is provided in Appendix B.

To conclude, after further leveraging the knowledge of company experts, the final pool of flow efficiency drivers is still subject to some adjustments. It is recognized that certain subtleties affecting SC flow efficiency are not fully captured during the literature research. Therefore, the following modifications are implemented:

- Number of production steps: Introduced as a flow efficiency driver. Since the analysis focuses solely on in-house production products and this parameter integrates the flow efficiency calculation formula, it is introduced as an explanatory variable.
- Unit production cost breakdown: Previously identified as a flow efficiency driver, the unit production cost is separated into Material Cost, Production Cost, Material Overhead Cost, and Production Overhead Cost. This detailed breakdown is believed to better capture the relationship between different cost components and help identify processes' inefficiencies.
- Inventory replaced by safety stock: Inventory is replaced by safety stock due to the former being an extremely dynamic parameter constantly changing. Determining the precise moment for data extraction would be challenging. Conversely, safety stock is a more stable value with less time dependence.
- Response time replaced by machine time: Since manufacturing efficiency is being used as a proxy for SC flow efficiency, replacing response time with machine time makes sense, thereby focusing on the actual time considered relevant in this context.

The final pool of flow efficiency drivers is presented in Table 4.1.

Table 4.1: Final pool of flow efficiency drivers.

Driver	Driver
Annual demand	Number of production steps
Batch size	Product complexity
Forecast error	Production cost
Lead time	Production overhead cost
Machine time	Product value density (PVD)
Machine cost	Safety stock
Material overhead cost	Volume
Minimum order quantity (MOQ)	Weight

Once the desired data is identified, the extraction process involves the execution of specific queries in SAP Analysis for Microsoft Excel.

4.3 Data Preparation

Conducting data evaluation and preparation tasks before developing ML models is essential to ensure reliable outcomes (He et al., 2016). These preliminary steps help in identifying and rectifying errors and inconsistencies in the data, thereby enhancing its quality. Moreover, it is fundamental to structure the data in a way that makes it suitable for analysis, which includes normalizing values, encoding categorical variables, and addressing multicollinearity. This process not only improves the accuracy of the models but also increases their efficiency and stability.

4.3.1 Data Validation

The validity of the generated model is related to the underlying data. Thus, checking the quality of the data for the efficiency drivers is essential to ensure good validity of the model (VDI, 2015). Additionally, it is necessary to validate the data quality outputs obtained through the flow efficiency drivers evaluation matrix. According to VDI (2015), to ensure high quality of the data collected, the following points must be observed:

- **Error-free data - semantic:** It must be checked whether the data are correct with regard to their meaning. A value which initially appears numerical may well have a non-numerical meaning.
- **Error-free data - syntactical:** Consistent use of separating characters such as period or comma as number format or decimal separation must be checked. Furthermore, uniform dimensions must be used to ensure comparability between different products. Last, a consistent notation is essential to ensure that the model can run correctly.
- **Completeness of data:** In data sets, it can happen that data is completely missing or duplicates are present. If data or information is missing, it must either be corrected or the product must be excluded from the model. Duplicates can be prevented by giving each product a unique identifier.

- **Objectivity of the data:** In principle, the data collected should not depend on the judgment of the person who collected it. If this is the case, the reproducibility of the data is questionable.

4.3.2 Data Processing

Considering the data quality criteria defined in the last subsection as well as ML model requirements, some preparation tasks were performed to ensure that the model's development could be started:

- **Data cleaning:**
  - The flow efficiency drivers identification step deemed "Weight" as an indispensable parameter in the SC flow efficiency estimation. Thus, it is decided to leave out of the analysis all the materials for which there is no weight recorded. The same rationale is applied to materials in which the "Demand" parameter is missing. This step leads to the exclusion of 4,509 data points, with the final dataset used to feed the ML models, having the information of 2,076 different materials. This step resulted in the exclusion of 4,509 data points, leaving a final dataset of 2,076 materials for the ML models. Figure 4.2 summarizes the various filtering steps from the initial to the final dataset.
  - Some data points presented no value for the "Safety Stock" parameter. In these, the null value is attributed, as it is the system default.

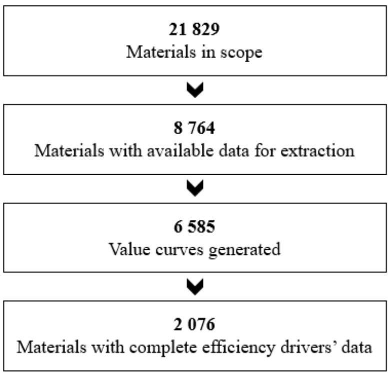


Figure 4.2: Filtering steps.

- **Data transformation:** The "Forecast Error" is calculated as the percentual difference between the forecasted demand and the actual demand for the same materials during 2023.

- **Feature Selection:** A Spearman correlation matrix is used to tackle the existence of multicollinearity in the flow efficiency drivers dataset, through the definition of a threshold from -0.85 to 0.85. The attributes with a higher positive or negative correlation between each other are eliminated by keeping the rest without having any further filtering application to have a variance in the data set.

Production overhead and material overhead costs are excluded from the analysis, and the data is eliminated from the dataset. No other dimensionality reduction techniques are applied since maintaining the interpretability of results is one of the great concerns of this methodological development. The final pool of flow efficiency drivers and their description is detailed in Appendix C.

- **Data Normalization:** To enhance the accuracy of ML models and prepare for subsequent clustering techniques, data normalization is employed. Normalization ensures that all attributes have the same weight in the analysis, a fundamental requirement for many techniques. The min-max normalization method is utilized, which scales the features to a fixed range (typically between 0 and 1) by subtracting the minimum value and dividing by the range of each attribute. This process ensures that all features contribute equally to the analysis, preventing attributes with larger scales from dominating the results. The formula for min-max normalization is as follows:

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{4.1}$$

- **Data Consolidation:** To integrate the flow efficiency values calculated through the value curves with the flow efficiency drivers dataset, a consolidation process is undertaken. This consolidation ensures that the dataset contains both the explanatory variables (flow efficiency drivers) and the target variable (flow efficiency values). By merging these datasets based on common identifiers, each observation in the dataset now includes both the predictors and the corresponding flow efficiency values, providing the necessary input for the training of ML models.

Chapter 5

Feature Selection for Supply Chain Flow Efficiency Prediction

This chapter offers a comprehensive overview of the methods employed to predict flow efficiency, building on the previously identified drivers. The approach aims to achieve an optimal balance between model complexity and performance. To this end, SBE is used to identify a smaller set of flow efficiency drivers that consistently deliver robust model performance (Khaire and Dhanalakshmi, 2022).

The first subsection outlines the criteria used to evaluate the models throughout the methodological development. Following this, the models and the corresponding results are presented.

5.1 Model's Evaluation

In predictive modelling, having a robust system for model evaluation is crucial. This importance is magnified when multiple metrics are used, allowing for a more comprehensive assessment, and capturing different aspects of the model's accuracy and reliability.

In this context, R^2 (Coefficient of Determination) and RMSE (Root Mean Squared Error) are used as the model's evaluation metrics. By doing so, a balanced evaluation is ensured as this dual-metric approach helps identify models that fit the training data well and perform consistently on new, unseen data.

The process begins with the initial selection of the model with the highest R^2 value, ensuring a strong fit to the data. This initial model serves as a starting point for further comparisons.

Next, the marginal differences between the normalized RMSE and R^2 values are calculated for all the trained models. This involves determining how much each model improves or deteriorates in terms of RMSE and R^2 compared to the initially selected model. The goal is to find models that significantly improve one metric without disproportionately worsening the other.

The selection is then iteratively updated based on these marginal differences until no better trade-off is found, ensuring the chosen model provides the optimal balance between RMSE and R^2. By following this systematic approach, the evaluation system ensures that the final selected model is not only accurate but also generalizes well, achieving a balanced performance across both metrics.

5.2 Backward Stepwise Elimination

In this study, a diverse set of ML models is employed to predict the flow efficiency of the SC. Firstly, the selection of these models is guided by the unique characteristics of the dataset. Specifically, PLS regression is chosen due to the high degree of multicollinearity present in the dataset, while Beta regression is selected because the target variable is bounded between 0 and 1.

Additionally, a range of commonly used ML models are incorporated to provide a comprehensive analysis, leveraging their unique strengths. This strategic selection was informed by an in-depth analysis of the dataset to identify models with the potential for high predictive accuracy. The full list of models used is presented in Table 5.1.

Table 5.1: Machine learning models used for flow efficiency prediction.

Model	**Model**
AdaBoost	Neural Networks Regression
Bayesian Regression	Ordinary Least Squares (OLS) Regression
Beta Regression	Partial Least Squares (PLS) Regression
Decision Trees	Random Forest
K-Nearest Neighbors (K-NN)	XGBoost

All models are trained using k-fold cross-validation to ensure reliability and minimize the risk of overfitting. Additionally, each model is fine-tuned by optimizing their specific hyperparameters to achieve optimal performance. A dataset comprising information on 1,874 different materials is used to train the models. The performance results[1] are shown in Table 5.2.

Table 5.2: Machine learning models' flow efficiency prediction performance.

Model	**R^2**	**RMSE**
AdaBoost	0.684	0.069
Beta regression	0.396	0.106
Bayesian regression	0.274	0.108
Decision trees	0.803	0.053
K-nn	0.734	0.0614
OLS	0.219	0.110
PLS	0.213	0.110
Random forest	0.876	0.043
XGBoost	0.899	0.034

The results highlight that non-linear and ensemble methods, particularly XGBoost and Random Forest, significantly outperform linear methods in predicting flow efficiency. These models are

[1] Neural Networks is excluded from the analysis as the models don't yield a positive value for R^2

better suited to capture the complex relationships in the dataset. Linear methods like OLS, PLS and Bayesian regression perform poorly.

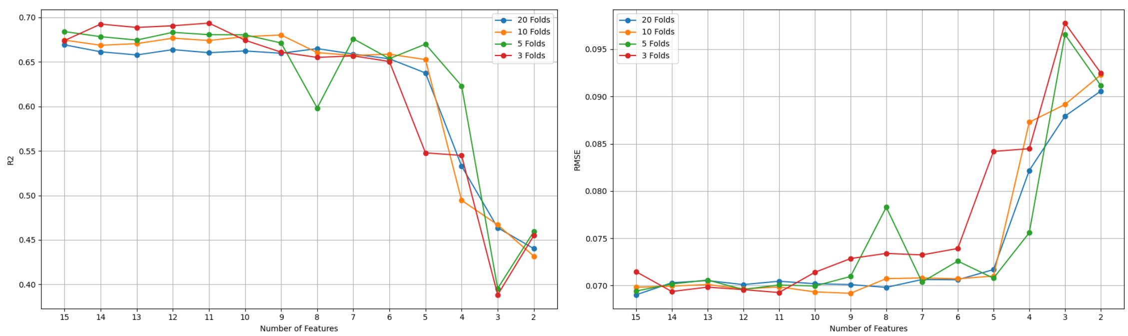
The models are subsequently run using SBE to further enhance predictive performance and understand the significance of different features. The feature importance metric in this method is calculated differently for various models.

For ensemble methods like Random Forest, XGBoost, AdaBoost, and Decision Trees, feature importance is determined based on each feature's contribution to the reduction in impurity—measured by Gini impurity or entropy—across all trees. In contrast, linear models such as OLS and Bayesian regression assess feature importance by the absolute values of their model parameters. For PLS regression, the importance is derived from the absolute values of the regression parameters for each feature.

By employing this approach, the study aimed to balance model simplicity and predictive accuracy, ensuring that the essential patterns in the dataset are captured effectively while maintaining computational efficiency and interpretability.

The output of this process included R^2 and RMSE values for each model across every k-fold cross-validation and for each set of features, sequentially extracted in order of importance until only one feature remained. The detailed results illustrating the impact of feature reduction on model performance can be found in the Appendix D.

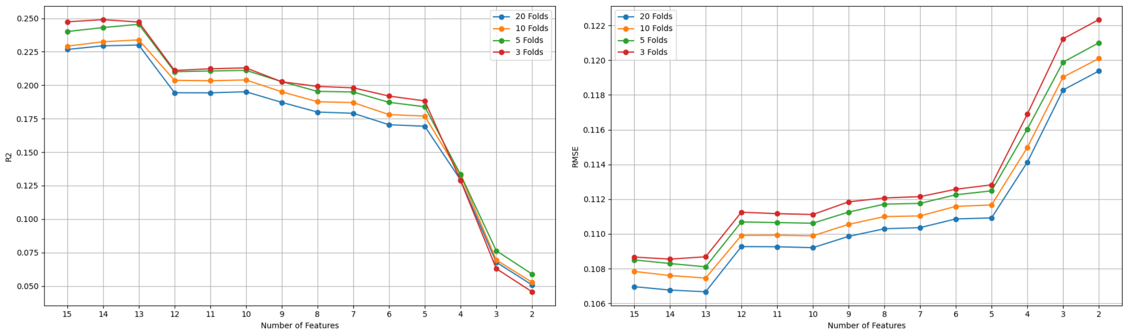
The plots[2] presented from 5.1 to 5.9 describe the evolution of the model's performance with dimensionality reduction.



(a) R^2 for AdaBoost (b) RMSE for AdaBoost

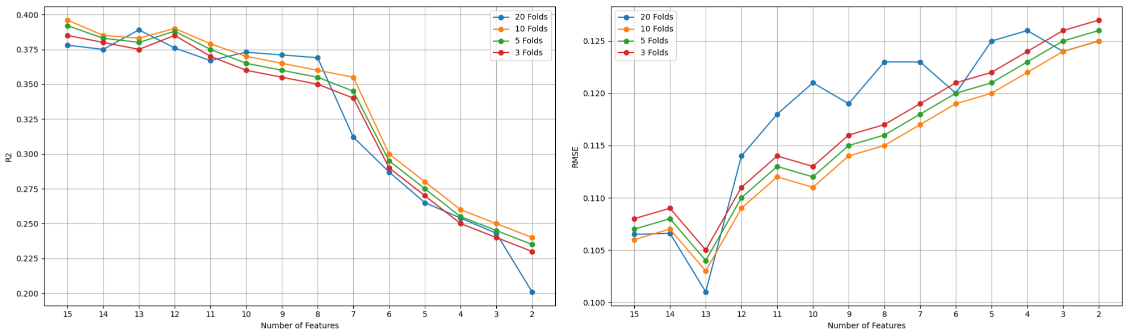
Figure 5.1: AdaBoost model complexity vs model performance.

[2]Neural Networks is excluded from the analysis as the models did not yield a positive value for R^2



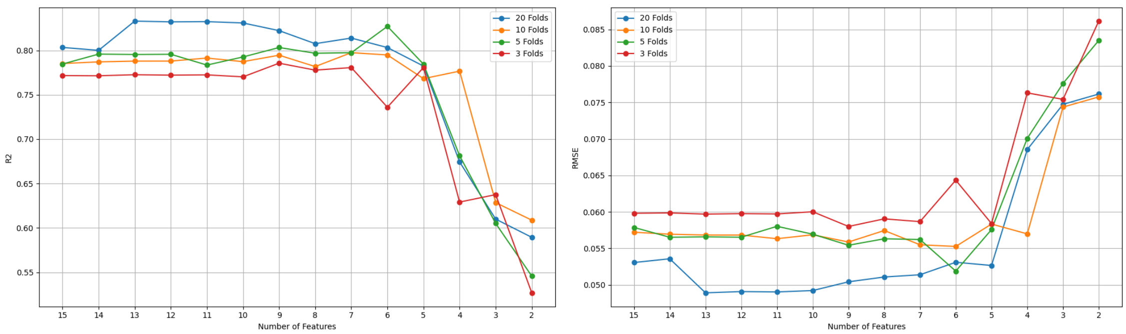
(a) R² for Bayesian Regression (b) RMSE for Bayesian Regression

Figure 5.2: Bayesian Regression model complexity vs model performance.



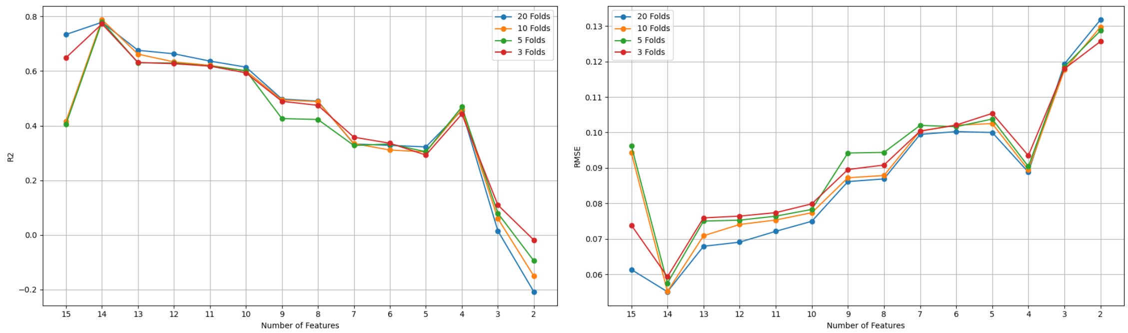
(a) R² for Beta Regression (b) RMSE for Beta Regression

Figure 5.3: Beta Regression model complexity vs model performance.



(a) R² for Decision Trees (b) RMSE for Decision Trees

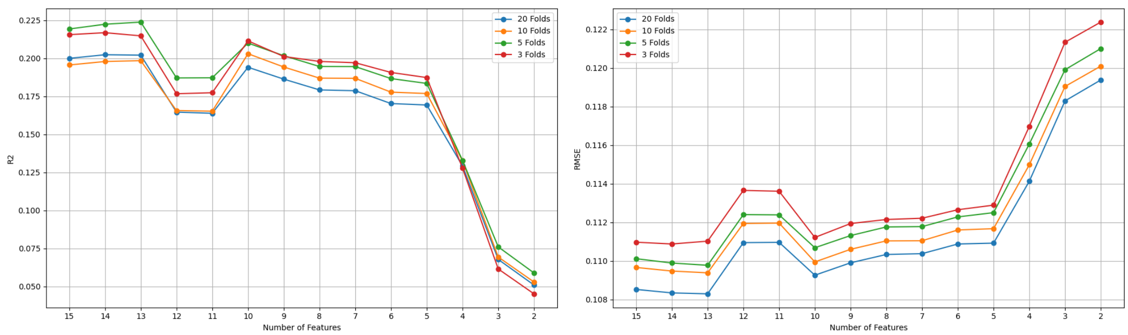
Figure 5.4: Decision Trees model complexity vs model performance.



(a) R² for K-nn

(b) RMSE for K-nn

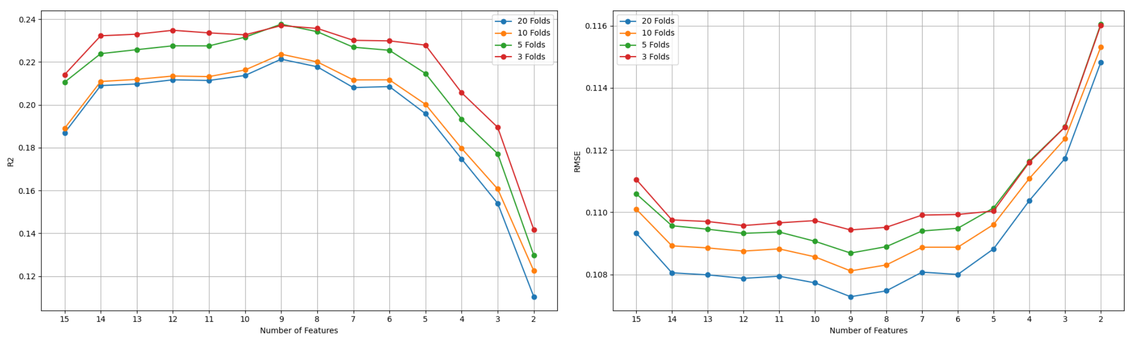
Figure 5.5: K-nn model complexity vs model performance.



(a) R² for OLS

(b) RMSE for OLS

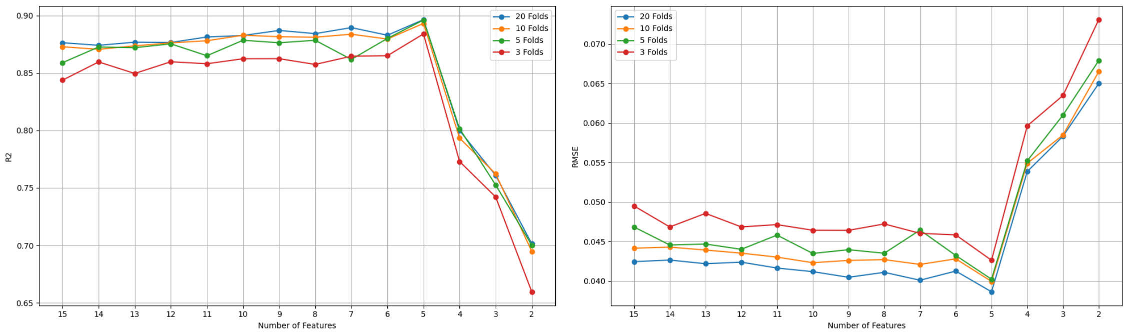
Figure 5.6: OLS model complexity vs model performance.



(a) R² for PLS

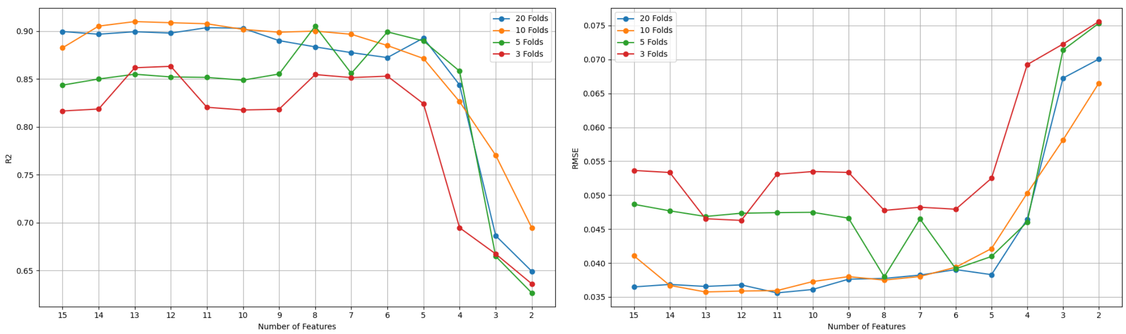
(b) RMSE for PLS

Figure 5.7: PLS model complexity vs model performance.



(a) R² for Random Forest (b) RMSE for Random Forest

Figure 5.8: Random Forest model complexity vs model performance.



(a) R² for XGBoost (b) RMSE for XGBoost

Figure 5.9: XGBoost model complexity vs model performance.

Based on the analysis of the plots, Decision Trees, Random Forest, and XGBoost have been selected for further analysis due to their robust performance during dimensionality reduction. It is observed that the performance of the three models remains stable until the 8^{th} feature is excluded. The optimal number of folds for each of the best-performing models was determined using the evaluation system described in the previous section. The results indicated that K=20 is optimal for Decision Trees and XGBoost, while K=10 is optimal for Random Forest For these specific parameters, the models maintain R^2 values above 0.80 and RMSE values below 0.053, even when more than 50% of the features are excluded from the training dataset.

The subsequent analysis aims to identify the final set of features used in each of these best-performing models before their performance metrics started to decline. The set of features may differ across models because the underlying algorithms assign different importance coefficients to various features. Consequently, the order in which features are excluded is not the same for each model, as described in Appendix E.

The features used for model training before the decline in performance metrics are listed in Table 5.3. The decline moment corresponds to the 8th iteration, during which the 8th least important feature was excluded. This table ranks the features by their importance, reflecting the order of exclusion until only one feature remains, for each one of the best-performing models. Additionally, the comparison of model performance before and after dimensionality reduction is presented in Table 5.4.

Table 5.3: Most important features identified through stepwise backward elimination.

Iteration	Decision Tree	Random Forest	XGBoost
8	lead time	lead time	machine time
9	#steps	#steps	material cost
10	material cost	bom	#steps
11	bom	machine time	lead time
12	machine time	material cost	bom
13	volume	volume	pvd

The analysis of Table 5.3 reveals that most of the features consistently appear across the three models, except PVD and the volume. Their repeated importance across different models suggests that they capture fundamental aspects of the underlying patterns in the data, contributing significantly to the predictive accuracy and robustness of the models. Focusing on these features can improve model performance, enhance generalization, and ensure that the models are leveraging the most informative variables.

Table 5.4: Performance of best-performing machine learning models before and after dimensionality reduction.

Model	All features		Dimensionality reduction	
	R^2	RMSE	R^2	RMSE
Decision Trees	0.803	0.053	0.8138	0.053
Random Forest	0.873	0.044	0.879	0.042
XGBoost	0.899	0.036	0.893	0.0382

Furthermore, the examination of model performance before and after dimensionality reduction, as shown in Table 5.4, demonstrates that all three models—Decision Trees, Random Forest, and XGBoost relatively maintained their performance after reducing the number of features, indicating that dimensionality reduction effectively eliminated irrelevant features and allowed the models to concentrate on the most significant variables. This process contributes to the models' robustness and generability.

Chapter 6

Clustering for Feature Importance Visualization

This chapter explores a different approach to understanding flow efficiency by investigating how various features influence different subsets of materials. Unlike Chapter 5, in which dimensionality reduction is used to train the model by focusing on the most significant features across the entire dataset, this chapter aims to delve deeper into the specific impacts of features on subsets of data.

The first subsection describes the methods used to split the data into different subsets. Following this, feature importance is evaluated within these subsets, testing the significance of the differences found.

The benefits of not building upon the findings from dimensionality reduction are significant in this context. By avoiding an overreliance on the reduced feature set identified in Chapter 5, the risk of excluding features critical for specific data subsets is mitigated. This approach allows for a more granular understanding of how different features impact the prediction of flow efficiency for different materials and justifies the analysis of every feature available in the dataset that comes from the methodology described in Chapter 4.

Gaining visibility into the specific ways in which features influence predictions within various subsets is essential to creating actionable strategies tailored to each material subset's unique characteristics.

6.1 Clustering

Clustering is employed to partition the initial dataset that arises from data consolidation procedures. Initially, the use of material groups as labels for creating clusters is considered, but this approach was ultimately abandoned due to concerns about compromising the integrity of the results. Material groups are defined from the sales department's perspective, categorizing products based on their end-use for the final customer. However, clustering data according to this criterion may lead to the loss of valuable information, as it may not be the optimal method for analyzing the characteristics that influence SC efficiency. Materials from different material groups might exhibit similar performance patterns, which could be overlooked using this approach. Additionally, the uneven distribution of materials across the identified material groups poses a significant challenge for a consistent analysis.

To address this, DBScan is used to create clusters based on the density of data points, allowing for a more natural grouping of materials based on their inherent characteristics, rather than predefined labels. This method ensures that true similarities in the data are reflected, leading to more insightful analyses of feature importance across different subsets.

Initially, DBSCAN identifies nine clusters along with one additional cluster labelled as noise. It is observed that the noise points, which are labelled as -1 by DBSCAN, constitute about 20% of the total dataset. Given that the entire dataset comprises only 1854 points, this percentage of noise is considered too high.

To ensure that the noise points are not excluded from the analysis, they are reassigned to the nearest existing clusters. This is achieved by finding the nearest neighbours of each noise point among the non-noise data points and reassigning each noise point to the cluster of its nearest neighbour. This iterative process ensures that all data points are included in meaningful clusters.

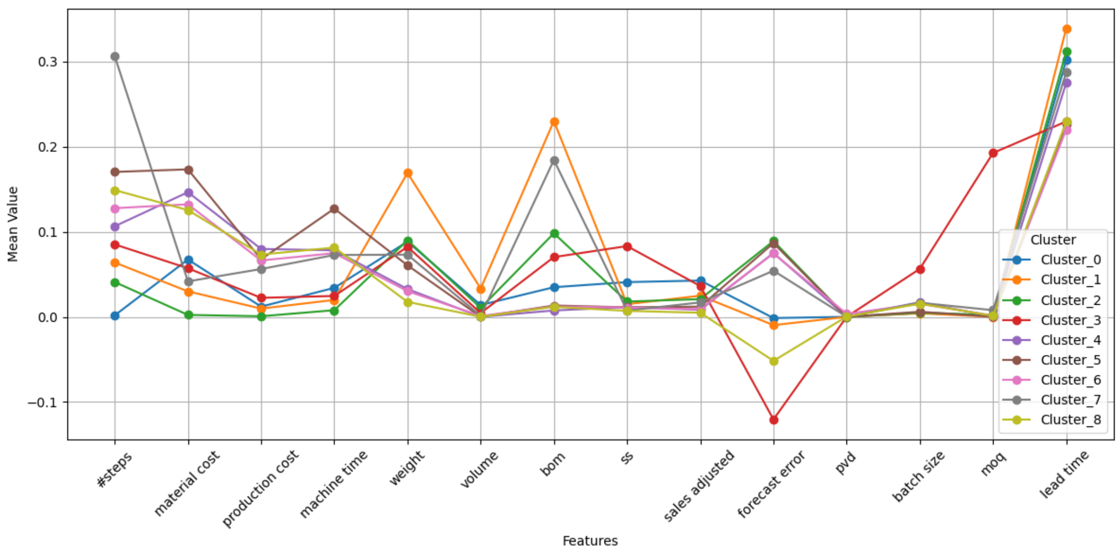


Figure 6.1: Profile cluster plot.

Table 6.1: Number of data points per cluster.

Cluster	Number of Data Points
0	74
1	232
2	103
3	88
4	197
5	240
6	381
7	124
8	426

The clusters identified are depicted in Figure 6.1. Each line in this profile cluster plot represents the average values of various features for a specific cluster. The x-axis corresponds to different features, while the y-axis shows their normalized mean values. This visualization allows for a comparative analysis of the clusters, highlighting their unique characteristics and differences. The distribution of data points among the clusters is detailed in Table 6.1.

Table 6.2: Material groups in each cluster.

Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
ph_GEEB	ph_GEEA	ph_GEEA	ph_GEEA	ph_GEEB	ph_GEDD	ph_GEEB	ph_GEEB	ph_GEEB
ph_GEDD	ph_GEEB	ph_GEEB	ph_GEEB	ph_GEDD	ph_GEEY	ph_GEDD	ph_GEDD	ph_GEEG
ph_GEEY	ph_GEDD	ph_GEEN	ph_GEEG	ph_GEEY		ph_GEEY	ph_GEDC	ph_GEDD
ph_GEDS	ph_GEDC	ph_GEDD	ph_GEDD				ph_GEEY	ph_GEEY
ph_GEMJ	ph_GEEY	ph_GEDC	ph_GEEY				ph_GEDG	
ph_GEDG	ph_GEDS	ph_GEEY	ph_GEDS				ph_GEDH	
ph_GEMO	ph_GEES	ph_GEES						
ph_GEDH		ph_GEMJ						
ph_GEMR		ph_GEDG						
		ph_GEMR						

Analysis of Table 6.2 reveals that materials within the same material group are not consistently clustered together. Different clusters contain materials from the same material group, indicating that clustering by the material group would obscure important data subtleties crucial for predicting flow efficiency.

6.2 Model's Development

After splitting the original dataset into clusters, each new dataset was trained using the models presented in Table 5.1 from Chapter 5.

Once again, the methodology for model development described in Chapter 5 is employed. Each model is trained using k-fold cross-validation and hyperparameter tuning to optimize performance.

The performance of each model is then evaluated for each cluster individually, using the evaluation system described in Section 5.1. The results for every combination between cluster and model are included in Appendix F[1], while the results of the best-performing model for each cluster are

[1]Combinations that don't yield a positive value for R^2 are excluded from the analysis.

presented in Table 6.3. Most models exhibit strong predictive power, with R^2 values exceeding 0.8 for seven out of the nine clusters, suggesting that the models explain a significant portion of the variance in efficiency. The AdaBoost and XGBoost models, in particular, demonstrate exceptional accuracy, achieving R^2 values above 0.9 and low RMSE values, indicative of minimal prediction errors. However, the KNN model for Cluster 7 shows comparatively lower performance with an R^2 of 0.573, highlighting areas where prediction accuracy could be improved.

Table 6.3: Performance of best-Performing models for each cluster.

Cluster	Model	R^2	RMSE
0	AdaBoost	0.734	0.112
1	AdaBoost	0.956	0.031
2	AdaBoost	0.806	0.097
3	AdaBoost	0.968	0.023
4	XGBoost	0.885	0.007
5	XGboost	0.940	0.012
6	XGBoost	0.899	0.017
7	KNN	0.573	0.054
8	Random Forest	0.887	0.010

Appendix G includes plots illustrating the differences between the actual efficiency values and the predicted values, for best best-performing model in each one of the clusters.

Comparing these results with those presented in Table 5.2, one can see that, except for Clusters 0 and 7, clustering significantly enhances the predictive accuracy of the models. This improvement can be attributed to the fact that clustering enables the models to capture the unique patterns within each subset of data. By focusing on more homogeneous groups, the models can achieve higher accuracy as they are tailored to specific nuances. However, it is also important to note that clustering can reduce the generalization capability of the models. When a model is tailored to specific clusters, its ability to generalize across the entire dataset may be diminished, as is portrayed in Clusters 0 and 7, where the predictive accuracy is lower, suggesting that these clusters might be more variable or not as well-defined as others.

6.3 Shap Values for Feature Importance

After identifying the best-performing model for each cluster, SHAP values are extracted for these specific combinations to analyze the impact of different features on the model's predictive ability.

No feature selection was applied, as maintaining granular interpretability is essential to draw conclusions about the hypothesis that different sets of materials are impacted differently by the identified flow efficiency drivers.

To ensure a reliable comparison of SHAP values across different models, the Kernel explainer is used (Stenwig et al., 2022). The Kernel explainer is a model-agnostic method that allows the calculation of SHAP values for any model, regardless of the underlying algorithm.

Figures 6.2 to 6.10 illustrate how the selected features affect the predictive ability of the best-performing model in each cluster. Beeswarm plots are used to provide clear visibility of these relationships.

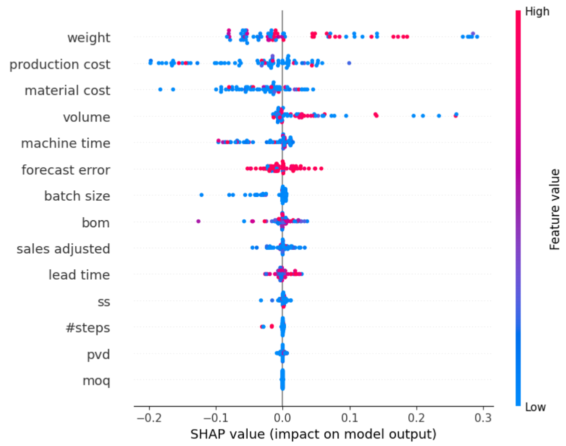


Figure 6.2: Shap values distribution for AdaBoost in cluster 0.

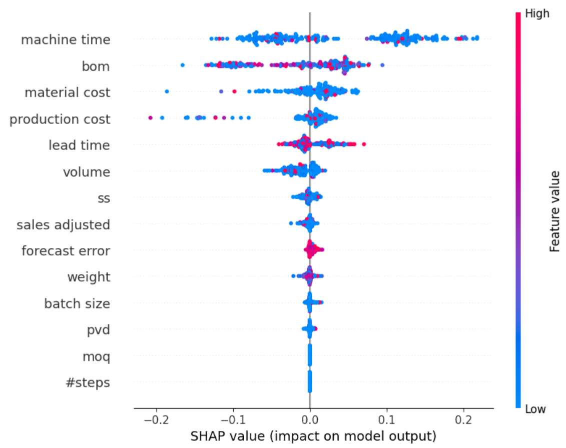


Figure 6.3: Shap values distribution for AdaBoost in cluster 1.

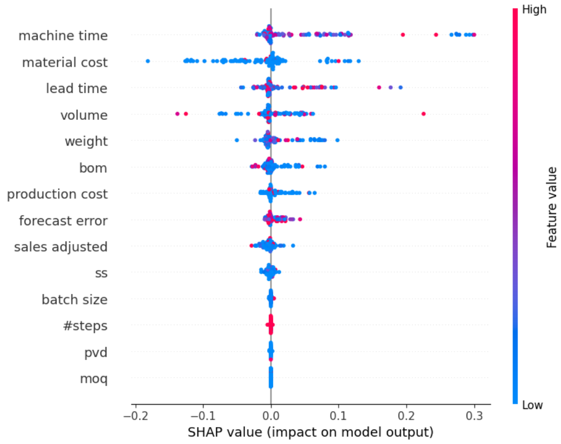


Figure 6.4: Shap values distribution for AdaBoost in cluster 2.

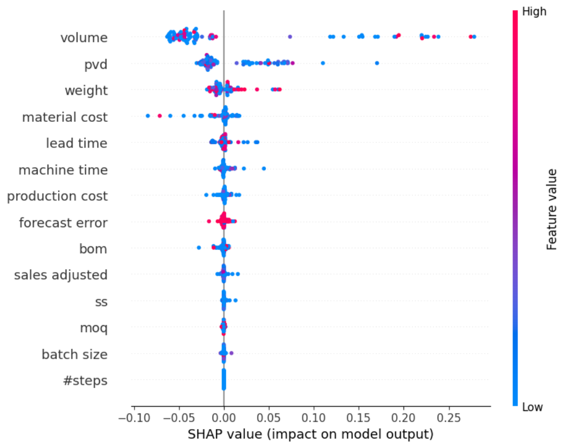


Figure 6.5: Shap values distribution for AdaBoost in cluster 3.

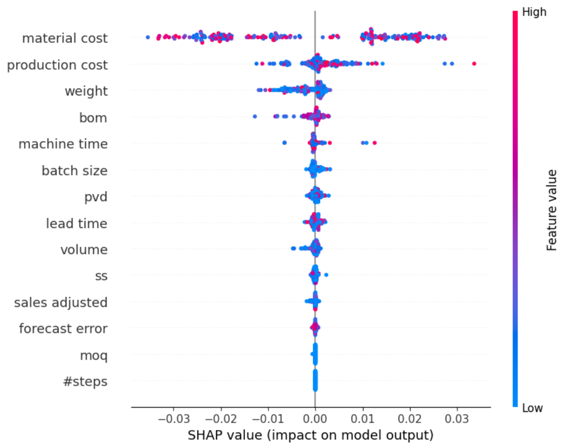


Figure 6.6: Shap values distribution for XGBoost in cluster 4.

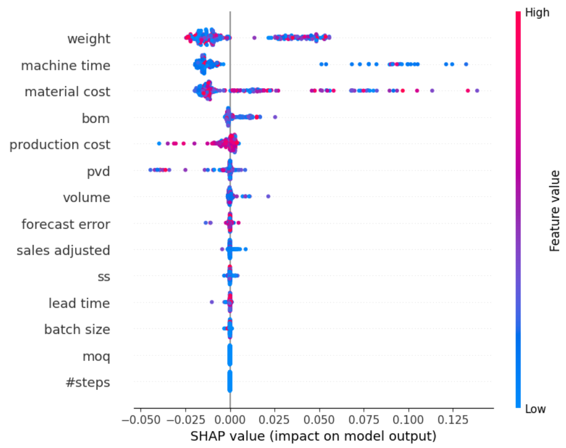


Figure 6.7: Shap values distribution for XGBoost in cluster 5.

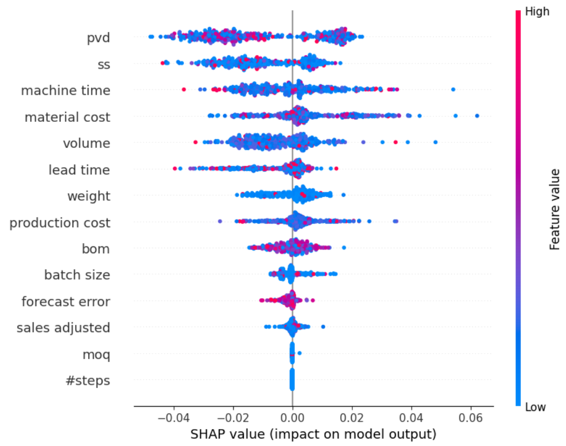


Figure 6.8: Shap values distribution for XGBoost in cluster 6.

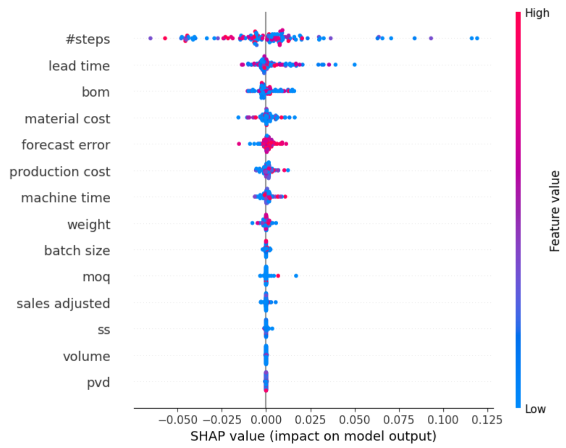


Figure 6.9: Shap values distribution for K-nn in cluster 7.

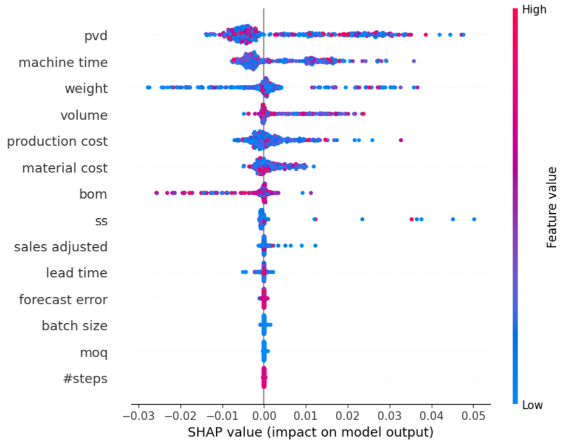


Figure 6.10: Shap values distribution for AdaBoost in cluster 8.

From the thorough analysis of the plots, three key observations emerged:

- **Feature Importance Variability:** The analysis of SHAP beeswarm plots reveals that different clusters have distinct sets of most impactful features. This variability underscores the hypothesis that different sets of materials have their flow efficiency being influenced in diverse ways by the flow efficiency drivers. Such differences highlight the necessity of considering cluster-specific characteristics when evaluating efficiency.
- **Consistent Features:** Despite the variability, certain features such as machine time and machine cost consistently appear as significant across multiple clusters, corroborating the results from the methodology described in Chapter 5. However, it is crucial to note that their impact on efficiency is not uniform. For example, material cost generally negatively impacts efficiency, but the degree of this impact varies from cluster to cluster. This consistency in feature appearance, combined with varying impacts, suggests that while some drivers are universally important, their influence depends on cluster-specific factors.
- **Impact of Features:** Detailed examination of the SHAP values indicates that features like the number of steps, lead time, and batch size often negatively impact efficiency when their values are high. Conversely, features such as weight and PVD can positively influence efficiency in certain clusters. These insights reveal the complex and sometimes counterintuitive nature of feature impacts on flow efficiency.

6.4 Statistical Validation

The SHAP beeswarm plots provide compelling evidence that, while certain features consistently emerge as important across different clusters, their impacts on the prediction of flow efficiency can vary significantly. However, it is crucial to statistically validate these findings to provide a higher degree of reliability (Guleria, 2024). If it is statistically proven that the differences between the SHAP values of the same features for different clusters are statistically significant, it is possible to validate the hypothesis that the clusters are indeed affected in different ways by the flow efficiency drivers.

To perform any kind of statistical evaluation is crucial to understand if the data under analysis is normally distributed as this characteristic has a direct impact in the typology of tests that can be carried out. To determine whether the data follows a normal distribution, the Shapiro-Wilk test is applied to the SHAP values of each feature within each cluster. This test is particularly suitable for small sample sizes and is a standard approach for assessing normality. The null hypothesis (H0) states that the data is normally distributed, while the alternative hypothesis (H1) states that the data is not. If the p-value from the Shapiro-Wilk test is greater than 0.05, we fail to reject the null hypothesis, indicating that the data is normally distributed. The results for this hypothesis test can be found in Appendix H.

Based on the results of the Shapiro-Wilk test, either an ANOVA test or a Kruskal-Wallis test is employed to assess the variance of SHAP values between clusters with the results being included in the Table 6.4. ANOVA is used when the SHAP values were normally distributed across clusters. The null hypothesis for ANOVA states that the means of the SHAP values are equal across clusters, while the alternative hypothesis suggests that at least one cluster has a different mean SHAP value. Conversely, the Kruskal-Wallis test was used when the SHAP values were not normally distributed. In this case, the null hypothesis states that the distributions of the SHAP values are equal across clusters, whereas the alternative hypothesis indicates that at least one cluster has a different distribution of SHAP values.

As in this case, the data for all the features in every cluster cannot be considered normally distributed, Krusdal-Wallis test is employed in every case.

Table 6.4: Kruskal-Wallis Test Results.

Feature	Statistic	P-Value
steps	46.240	2.140E-07
material cost	181.643	4.650E-35
production cost	126.379	1.591E-23
machine time	279.463	9.604E-56
weight	71.596	2.363E-12
volume	255.692	1.070E-50
bom	18.211	0.020
ss	309.612	3.700E-62
sales adjusted	88.764	8.285E-16
forecast error	303.058	9.201E-61
pvd	120.005	3.307E-22
batch size	189.504	1.036E-36
moq	36.327	1.530E-05
lead time	54.143	6.477E-09

The results from these statistical tests are presented in the form of test statistics and p-values. If the p-value is less than or equal to 0.05, it is concluded that there is a statistically significant difference in the impact of the feature between the clusters and the null hypothesis is rejected. Conversely, if the p-value is greater than 0.05, the null hypothesis cannot be rejected, indicating no significant difference in the feature's impact between the clusters. Through the analysis of Table 6.4, and taking into account how low is the p-value in the great majority of the test, the null hypothesis is rejected.

This rigorous approach ensures that the differences observed in the SHAP values are not merely by chance but are statistically significant, thus confirming that the clusters are indeed affected in different ways by the flow efficiency drivers. By validating these findings statistically, a higher degree of confidence in conclusions is provided, supporting the tailored strategies for improving flow efficiency based on the unique characteristics of each cluster.

Chapter 7

Case Study

In this chapter, the company profile is presented and the methodologies described in Chapter 5 and Chapter 6 are applied to a specific use case. By doing so, it is possible to illustrate how these methodologies can be effectively implemented in a practical setting.

7.1 Company Profile

Hilti, renowned for its cutting-edge products, systems, and services, is a global leader in the building and energy sectors. Operating across 120 countries, the company boasts a workforce of over 34,000 dedicated team members. With manufacturing facilities strategically located in Europe (including Liechtenstein, Austria, Germany, Hungary, and Norway), Asia (China, India, Malaysia, and Taiwan), and North America (Mexico and USA), Hilti ensures efficient production and distribution worldwide (Hilti, 2024).

Established in 1941 by Martin Hilti and his brother Eugen in Schaan, Liechtenstein, the company remains family-owned, with all shares held by the Martin Hilti Family Fund. Hilti's core values revolve around quality, innovation, and fostering close customer relationships.

In 2023 alone, Hilti achieved sales exceeding CHF 6.5 billion, a testament to its market prominence and customer trust. Operating on a direct sales model, approximately 25,000 employees engage directly with customers, ensuring personalized service and satisfaction (Hilti, 2024).

The company holds an extensive product portfolio comprising over 10,000 items, with 60 new product families introduced annually. This commitment to innovation drives Hilti's continuous evolution and reinforces its position as an industry pioneer.

The diagram in Figure 7.1 illustrates the relationship between the Sourcing Excellence department (DOS) and various business units (BU). Each BU operates independently and is responsible for its own performance. DOS plays a crucial role by providing support in the global procurement process,enhancing and digitalizing their operations. This strategic positioning allows DOS to accumulate expertise in SCM, given its interactions with diverse business practices. Furthermore, driving substantial changes that have a significant impact across the entire company.

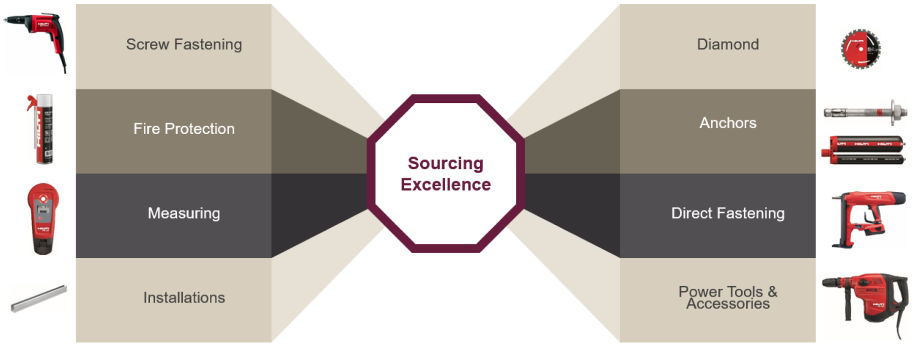


Figure 7.1: DOS as overarching cross-BU department to facilitate procurement process (Hilti, 2024).

7.2 Data Availability

In the past few years, Hilti has been adopting a global approach to its procurement process and, more broadly, to its SC. Contributing to this effort, is the global deployment of an ERP package meant to cover all day-to-day operations of the enterprise. This solution is completed with a database that stores all the information and makes it available to everyone with the required access across the company.

Despite the abundance of data available, the methodological development described in Chapters 6 and 7 is deeply related to the quality of the data. Therefore, while identifying the data in scope for this analysis, several constraints were taken into account.

The data chosen for this analysis encompassed all the business areas of general electric tools, which include three different BU: Power Tools & Accessories (PT&A), Measuring, and Diamond. The rationale behind this choice is explained by the fact that these units represent 70% of Hilti's total revenue, reflecting their significance (Hilti, 2024). Secondly, the greatest share of Hilti's in-house production also originates from these three BU. The in-house production criterion is crucial to ensure that the materials can be used in this analysis, as data from the production steps is needed for drafting the value curves, which are fundamental to the methodological development.

As previously mentioned, 2076 different materials have all the necessary information available for this analysis, encompassing 16 different material groups produced in five different Hilti plants around the globe. The diversity of this data is a key concern while defining the materials in scope, as it is an essential criterion for building a robust model capable of performing accurately on unseen data.

From the total dataset, 1864 materials are used for model training, while 210 materials are purposely left out to test the model's performance on unseen data. The results of the methodology described in Chapters 6 and 7, applied to this specific set, form the main focus of this case study.

These 210 materials represent a specific material group: rotary hammers, which are depicted in Figure 7.2. Rotary hammers are power tools designed primarily for heavy-duty drilling and chiselling applications in construction and industrial settings. Known for their durability and efficiency, Hilti rotary hammers combine rotation with a powerful hammering action, making them ideal for drilling through tough materials such as concrete, brick, and masonry.

Figure 7.2: NURON TE 4-22 Cordless Rotary Hammer SDS Plus (Hilti, 2024).

The selection of the rotary hammer material group is deliberate, as it is one of Hilti's best-selling power tools. Evaluating the efficiency of the SC for this product group and identifying possible areas of improvement makes strategic sense, given its significant impact on the BU PT&A performance.

7.3 Flow Efficiency Prediction through Feature Selection Stepwise Approach

Building on the conclusions drawn throughout Chapter 5, the data on rotary hammers is used as a testing set for the best-performing models identified earlier. The first prediction iteration includes all 14 different flow efficiency drivers. Subsequently, the models are trained and tested with only the 7 most impactful features, as identified through SBE, included in Table 7.1

Table 7.1: Most important features identified in Chapter 5 after stepwise backwards elimination.

Feature
Bom
Lead Time
Machine Time
Material cost
Number of steps
PVD
Volume

Table 7.2: Performance of best-performing machine learning models before and after dimensionality reduction on unseen data

Model	All features		Dimensionality reduction	
	R^2	RMSE	R^2	RMSE
Decision Trees	0.401	0.174	0.321	0.185
Random Forest	0.521	0.155	0.395	0.175
XGBoost	0.571	0.147	0.596	0.143

The use of unseen data aims to validate the conclusions from Chapter 5. The results from testing on unseen data reveal a significant drop in model performance when compared to the results that come from cross-validation on the training set, included in Table 5.4. Specifically, while the models performed with R^2 values of 0.899, 0.873, and 0.803 for XGBoost, Random Forest, and Decision Trees, respectively, during training with all features, their performance on unseen data fell to 0.596, 0.521, and 0.401, respectively. This substantial decline highlights a crucial challenge in model generalization.

Two main factors could be contributing to this performance gap:

- **Overfitting during training:** Theoretically, the high performance on the training set, especially with all features, may suggest that the models might have overfitted. This phenomenon occurs when a model learns the noise and specific patterns in the training set rather than the underlying relationship, leading to poor performance on unseen data. However, cross-validation is used throughout all the training processes, making this hypothesis less likely to be valid.

- **Data variability and quality:** The unseen data may have characteristics or variability not present in the training set. Clear differences in data patterns can affect model performance, as models trained on one distribution might fail to generalize to another.

Moreover, through the evaluation of results before and after dimensionality reduction, a notable performance difference is also identified. During training, models with dimensionality reduction either maintained or improved their prediction ability, as evidenced by their R^2 values, in Table 5.4. However, this trend does not hold for every model tested for unseen data. Decision Trees and Random Forests with dimensionality reduction perform worse when compared to those using all features. As shown in Table 7.2, XGBoost is the only model presenting a slight improvement.

This variance in performance on unseen data may suggest that the selected features during dimensionality reduction might not adequately capture the data patterns present in the new, unseen dataset. While dimensionality reduction helped to avoid overfitting during training, it appears that the reduced feature set was less effective at generalizing to unseen variations while using Decision Trees and Random Forests. This discrepancy indicates that some relevant features may have been excluded during the reduction process, highlighting the need for a more robust feature selection method that can better capture the essential characteristics of both the training and unseen data.

7.3.1 Performance Pricing in Flow Efficiency

Hilti's supply teams frequently utilize Performance Pricing as a method for price comparison. This analysis adopts a similar approach to make Flow Efficiency across different products comparable. This method plots theoretical Flow Efficiency values on the y-axis and predicted values on the x-axis.

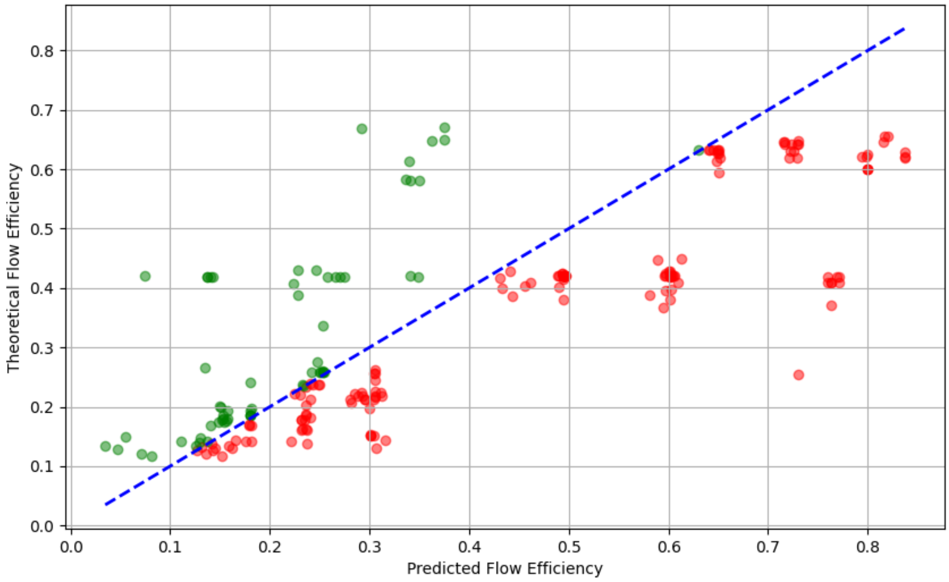


Figure 7.3: Application of performance pricing to flow efficiency.

Figure 7.3 illustrates the results of this approach, using the outputs of XGBoost model after dimensionality reduction as it was the one which provided the better and most consistent results. Products represented in green have theoretical Flow Efficiency values higher than those predicted by the XGBoost model, indicating favourable outcomes. On the other hand, products depicted in red have theoretical Flow Efficiency values lower than the predicted ones, which are less desirable.

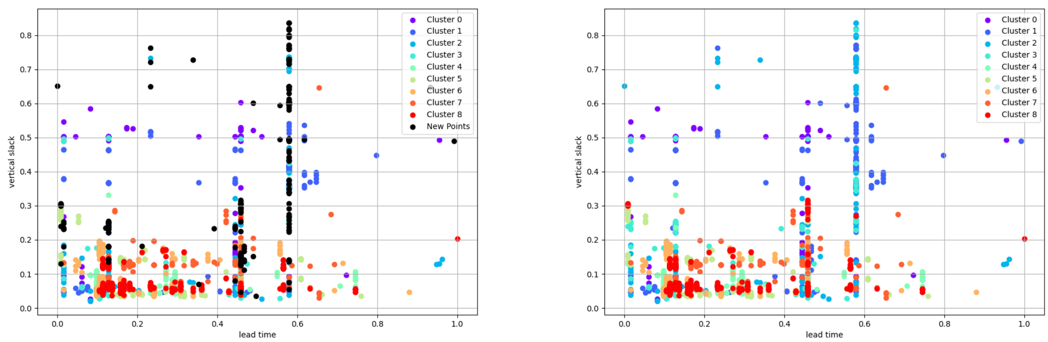
Products lying along the 45-degree line, known as the average line, have identical theoretical and predicted Flow Efficiency values, marking them as favourable. The interpretation of the model's outcomes is based on their placement relative to this line.

Products shown in green indicate higher theoretical Flow Efficiency than predicted, signifying optimal efficiency and suggesting that these products do not require immediate efficiency improvement efforts. Conversely, products in red highlight a discrepancy where theoretical Flow Efficiency is lower than predicted, indicating inefficiencies and room for improvements to bring their efficiency closer to the average line. Products positioned on the average line demonstrate an alignment between theoretical and predicted efficiencies. Although these products are performing as expected, they still present opportunities for further efficiency enhancements to maintain or improve performance.

In summary, by employing the Performance Pricing approach, it is possible not only to effectively identify areas where Hilti's SC can be optimized for better Flow Efficiency but also to define a priority ranking by considering the gap difference in efficiency.

7.4 Clustering for Flow Efficiency Prediction and Features Importance Visualization

Building upon the findings of Chapter 6, the new data points were assigned to one of the pre-existing clusters, through K-nn method. The distance to the central point is calculated and the points are assigned to the nearest cluster. The new points are shown in the plots of Figures 7.4a and 7.4b.



(a) New data points

(b) Final clusters after new points allocation

Figure 7.4: Unseen data allocation to clusters.

The distribution of the new points across the clusters is described in Table 7.3. It is noteworthy that unseen data from one specific material group is distributed across seven different clusters. This finding corroborates the conclusions drawn in Chapter 6, which highlight that even within the same material group, flow efficiency drivers impact efficiency in diverse ways.

Table 7.3: New data points

Cluster	New Data Points	Final Cluster size
0	0	74
1	78	310
2	49	152
3	43	131
4	3	200
5	3	243
6	0	381
7	10	134
8	24	450

Taking into consideration the best model-cluster combination for predicting flow efficiency identified in Chapter 6, flow efficiency is predicted for each cluster individually, using the new data as test data, which was not used during model training. For each cluster, the respective best model is utilized to make predictions, ensuring that the models are appropriately tailored to the unique characteristics of each cluster. After obtaining the predictions from the models across the 7 clusters under analysis, the overall R^2 and RMSE are calculated to assess the performance of the predictions. The results yielded an R^2 of 0.558 and an RMSE of 0.160, indicating similar results to the ones yielded throughout the methodology included in Table 7.2.

Chapter 8

Discussion and Outlook

In this concluding chapter, the contributions of this dissertation to both practice and science are discussed. The limitations of the research conducted are detailed, providing a foundation to support recommendations for future research.

8.1 Results

The proposed master thesis, "Supply Chain Efficiency Prediction: Leveraging Machine Learning for Improved Accuracy and Interpretability" aims to develop innovative methodologies to predict flow efficiency more precisely while offering clear insights into the features that most significantly impact these predictions, leveraging ML techniques.

Firstly, flow efficiency values for various Hilti products are predicted using 14 different explanatory variables through supervised learning techniques. Two approaches are used to achieve these results. The first approach involves analysing the entire dataset as a whole, with the best results obtained from the XGBoost model, achieving an R^2 of 0.571 and an RMSE of 0.147. In the second approach, the initial data set is divided into several clusters based on inherent data patterns, and predictions are made at the cluster level. The best-performing models for these clusters include AdaBoost, XGBoost, K-Nearest Neighbours (K-NN) and Random Forest, which yields combined results of 0.588 for R^2 and 0.160 for RMSE. However, for the sole purpose of flow efficiency prediction, the benefits of clustering the data are difficult to justify, as the improvements in prediction accuracy are marginal compared to the overall data analysis.

Secondly, a two-fold analysis is conducted to evaluate the existence of a set of features that consistently emerge as the most important to predict flow efficiency. In the first analysis, the entire dataset is used to apply SBE to identify the most significant features by evaluating the models' performance during the dimensionality reduction process. In both training and testing, the best model performance remains stable until one of the following features is removed: machine time, material cost, number of steps, lead time, BoM, PVD, and volume. In parallel, SHAP values are employed to assess feature importance at the cluster level. Features such as machine time and material cost are systematically identified as the most critical for flow efficiency prediction, confirming their consistent significance for flow efficiency prediction.

Thirdly, calculating SHAP values for the best-performing models within each identified cluster reveals that despite machine time and material cost consistently emerging as two of the most important features for flow efficiency prediction, different clusters have their flow efficiency impacted by different drivers. In fact, even when the most important features are the same, their impact can vary between positive and negative values. This variability supports the previously formulated hypothesis that different sets of materials have their flow efficiency influenced by different drivers and in different ways. These differences underscore the need to consider cluster-specific characteristics when evaluating efficiency and, more importantly, when defining target improvements for underperforming products. Although clustering does not significantly improve the predictive accuracy for flow efficiency, it significantly improves model interpretability, unveiling data patterns that remain hidden when evaluating the dataset as a whole.

On a side note, the clustering of the data into multiple subsets introduces a new question. Upon evaluating the clusters, it becomes evident that there is no similarity between the data distribution within the clusters and the pre-defined material groups. Materials from the same group are allocated to different clusters, and each cluster contains multiple material groups. This heterogeneity in the data distribution suggests that material group definitions based on product functionality may not be the optimal way to evaluate and study flow efficiency. This observation builds ground for further research, which will be discussed in the following sections.

8.2 Contribution to Practice

The objective of this master's thesis is to achieve a robust and reliable estimation of flow efficiency using ML techniques. While an accurate estimation is advantageous for professionals, the true value lies in analysing the actual and predicted flow efficiencies together, enabling quick and actionable identification of improvement opportunities in production setups: from the flow efficiency graph, the improvement potential for a specific item can be derived and related to the improvement potential of other items. Although identifying appropriate ML techniques was the primary aim of this thesis, the findings already have practical applications for practitioners at Hilti's plants. They can now efficiently assess their production setups and have already initiated the first improvement steps, such as modifying supplier lead times, which highlights the relevance and applicability of the research outcomes.

8.3 Contribution to Science

Due to the novelty of flow efficiency research, the body of existing literature is minimal. Considering the other Hilti-related theses on the topic, the main contribution lies in the methodical rigour and the data scope regarding the selection of suitable ML techniques. Whereas in previous theses, PLS was used with a rather limited set of items, the significantly extended data scope (starting with value curves for 8200 items extracted from the company‘s ERP system) allowed for a thorough examination of appropriate estimation techniques.

8.4 Limitations

The value curves used in the presented methodological development are only a proxy depiction of the overall SC, as they include only the production stages. They exclude all processes before raw materials reach the production sites and the processes until the products reach the final customer. As previously described, these value curves are used to calculate the target variable, which means they only reflect production efficiency. However, the flow efficiency drivers used as explanatory variables to predict this efficiency encompass features that reflect the entire behaviour of the SC, such as lead time. This discrepancy may result in less accurate predictions and reduced overall effectiveness of the models, as they do not fully capture the complexity and full scope of the SC processes.

One significant challenge encountered in this research is related to the data collection and consolidation process for flow efficiency drivers. Despite extensive efforts, it was not possible to replicate the comprehensive pool of data identified during the literature review. This shortfall is primarily attributed to data quality issues, such as the unavailability of many desired data points and inconsistencies in the data obtained. These data quality issues likely have a direct impact on the performance of the ML models, as some important variables for predicting flow efficiency may have been left out. The significant discrepancy observed between training performance and testing on unseen data, despite using cross-validation, further corroborates the limitations inherent to the data itself. This discrepancy highlights the challenges posed by incomplete data, underscoring the importance of comprehensive and high-quality data collection in the development of robust predictive models.

Moreover, the SBE approach used in this study involves starting with a full model and incrementally removing predictors based on specific criteria. However, this method does not explore all possible combinations of features, which means that it may overlook combinations that could enhance predictive accuracy.

It is also important to highlight a key limitation related to extracting actionable insights from SHAP values. Although SHAP values are effective in evaluating the model’s behaviour and highlighting which features the model considers important, they do not necessarily reflect true causal relationships. This can pose a significant challenge in translating SHAP value insights directly into practical actions, as the model’s importance rankings might not align with the real-world factors affecting flow efficiency, making it difficult to derive practical strategies based solely on SHAP values.

8.5 Avenues for Further Research

The findings of this master dissertation open up several avenues for future research aimed at enhancing the accuracy and interpretability of SC efficiency predictions. These potential approaches focus on improving data quality, exploring more sophisticated feature selection techniques, diversifying data samples, and further validating key hypotheses.

Firstly, the enhancement of data collection processes is crucial. Involving multidisciplinary teams in data collection can significantly improve data availability. These teams can bring diverse perspectives and expertise, facilitating cross-checks between different data sources and ensuring a more comprehensive and reliable dataset. Improved data availability will likely enhance the performance of ML models, leading to more accurate predictions of SC efficiency.

Secondly, the current SBE approach used for feature selection can be replaced with more advanced techniques. Recursive feature elimination (RFE) or ML-based approaches, such as feature importance ranking from ensemble methods, can be explored. These techniques can examine a broader range of feature interactions, potentially uncovering more complex relationships that stepwise methods may overlook. This broader exploration of feature interactions could lead to models with improved predictive power.

Furthermore, further research should be conducted to test the hypothesis that traditional material group classifications may not be the most effective way to study flow efficiency in the SC. The heterogeneity observed in data distribution across clusters, which did not align with pre-defined material groups, suggests that alternative classification methods might be more appropriate. Investigating different clustering algorithms and classification criteria could provide new insights into the drivers of flow efficiency, leading to more targeted and effective SC improvements.

Additionally, applying elastic net regression could further validate the importance of features identified in this study. Elastic net, which combines the properties of LASSO and Ridge regression, can handle multicollinearity and perform variable selection. This validation step is crucial to ensure the reliability of the identified most important features for predicting SC efficiency.

Moreover, the case study's approach to data sampling can be revisited. The study utilised data from a specific material group not included in the data used in model training, which, despite including material groups from the same product family, may have jeopardised the model's performance. Future research should consider sampling random data sets instead of relying on specific material groups.

Lastly, for practitioners, and as indicated by the experience with Performance Pricing, there is substantial potential for further research into enhancing measurement identification techniques. VDI (2015) clearly outlines the next steps needed to better analyze and draw conclusions from the flow efficiency graph, therefore it should be leveraged to provide valuable insights and pave the way for future advancements in this area.

Bibliography

A Charnes, W Cooper, A. Y. L. and Seiford, L. M. (1997). Data envelopment analysis theory, methodology and applications. *Journal of the Operational Research Society*, 48(3):332–333.

Abdi, H. and Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459.

Abonyi, J. and Feil, B. (2007). *Cluster analysis for data mining and system identification*. Springer Science & Business Media.

Aigner, D., C. A. L. P. S. (1977). Formulation and estimation of stochastic frontier production models. *Journal of Econometrics*, 21-37(6).

Aitken, J., Childerhouse, P., and Towill, D. (2003). The impact of product life cycle on supply chain strategy. *International Journal of Production Economics*, 85(2):127–140.

Alzubi, J., Nayyar, A., and Kumar, A. (2018). Machine learning from theory to algorithms: an overview. In *Journal of physics: conference series*, volume 1142, page 012012. IOP Publishing.

Ayesha, S., Hanif, M. K., and Talib, R. (2020). Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*, 59:44–58.

Barlow, H. B. (1989). Unsupervised learning. *Neural computation*, 1(3):295–311.

Bauer, P. W., Berger, A. N., Ferrier, G. D., and Humphrey, D. B. (1998). Consistency conditions for regulatory analysis of financial institutions: A comparison of frontier efficiency methods. *Journal of Economics and Business*, 50(2):85–114.

Bharadiya, J. (2023). Machine learning and ai in business intelligence: Trends and opportunities. *International Journal of Computer (IJC)*, pages 123–134.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer.

Bley, N. (2022). Supply Chain Efficiency: Entwicklung einer Methodik zur quantitativen Bewertung von Lieferketten am Beispiel der Hilti AG.

Bolón-Canedo, V., Sánchez-Maroño, N., and Alonso-Betanzos, A. (2016). Feature selection for high-dimensional data. *Progress in Artificial Intelligence*, 5:65–75.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta psychologica*, 81(3):211–241.

Burger, A. (2008). *Produktivität und Effizienz in Banken*.

Cai, J., Luo, J., Wang, S., and Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, 300:70–79.

Chan, F. (2003). Performance measurement in a supply chain. *International Journal of Advanced Manufacturing Technology*, 21(7):534 – 548. Cited by: 401.

Chandrashekar, G. and Sahin, F. (2014). A survey on feature selection methods. *Computers & electrical engineering*, 40(1):16–28.

Chatterjee, S. and Hadi, A. S. (2015). *Regression analysis by example*. John Wiley & Sons.

Cho, D. W., Lee, Y. H., Ahn, S. H., and Hwang, M. K. (2012). A framework for measuring the performance of service supply chain management. *Computers and Industrial Engineering*, 62(3):801 – 818. Cited by: 247.

Chopra, S. and Meindl, P. (2006). Supply chain performance: Achieving strategic fit and scope. *Supply Chain Management: Strategy, Planning, and Operations*, (2006):22–42.

Cooper, W. W., Seiford, L. M., and Zhu, J. (2003). Data Envelopment Analysis - History, Models and Interpretations. pages 1 – 39.

Craighead, C. W., Blackhurst, J., Rungtusanatham, M. J., and Handfield, R. B. (2007). The severity of supply chain disruptions: design characteristics and mitigation capabilities. *Decision sciences*, 38(1):131–156.

Danert, G. (1988). *Wirtschaftliche Gestaltung der Fertigungslogistik : Herrn Prof. Dr. Günter Danert zum 75. Geburtstag*. Stuttgart. Literaturangaben am Ende der einzelnen Kapitel Literaturangaben am Ende der einzelnen Kapitel.

Daraio, C. and Simar, L. (2007). *Advanced Robust and Nonparametric Methods in Efficiency Analysis, Methodology and Applications*. Studies in Productivity and Efficiency.

Deloof, M. (2003). Does working capital management affect profitability of belgian firms? *Journal of Business Finance & Accounting*, pages 573–587.

Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR)*, 27(3):326–327.

Dixit, K. K. and Gupta, H. (2023). Selection of industrial engineering programs using wsm method. In *2023 1st International Conference on Cognitive Computing and Engineering Education (ICCCEE)*, pages 1–5. IEEE.

Dyckhoff, H. (2003). *Grundzüge der Produktionswirtschaft, Einführung in die Theorie betrieblicher Wertschöpfung*.

Ertöz, L., Steinbach, M., and Kumar, V. (2003). Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data. In *Proceedings of the 2003 SIAM international conference on data mining*, pages 47–58. SIAM.

Farchi, C., Farchi, F., Touzi, B., and Mousrij, A. (2023). A sustainable performance assessment system for road freight transport based on artificial neural networks. *Ingenierie des Systemes d'Information*, 28(3):647.

Ferrari, S. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of applied statistics*, 31(7):799–815.

Fischer, H. and Fischer, H. (2011). *The central limit theorem from Laplace to Cauchy: changes in stochastic objectives and in analytical methods*. Springer.

Fisher, M. L. (1997). What is the right supply chain for your product? *Harvard Business Review*, pages 105–116.

Fisher, R. (2006). *Statistical Methods For Research Workers*. Cosmo study guides. Cosmo Publications.

Forrester, J. W. (1958). Industrial dynamics: A major breakthrough for decision makers. *Harvard Business Review*, 36:37–66.

García-Arca, J., González-Portela Garrido, A. T., and Prado-Prado, J. C. (2016). "packaging logistics" for improving performance in supply chains: the role of meta-standards implementation. *Production*, 26:261–272.

Ghaouta, A., Riad, M., and Okar, C. (2021). Machine learning for warehouse management: a conceptual framework. In *2021 Third International Conference on Transportation and Smart Technologies (TST)*, pages 30–35. IEEE.

Gomes, S. (2021). A Performance Pricing methodology based on capital efficiency to improve supply chain performance.

Guleria, P. (2024). Blending shapley values for feature ranking in machine learning: an analysis on educational data. *Neural Computing and Applications*, pages 1–25.

Gunasekaran, A., hung Lai, K., and Edwin Cheng, T. (2008). Responsive supply chain: A competitive strategy in a networked economy. *Omega*, 36(4):549–564. Special Issue on Logistics: New Perspectives and Challenges.

Halevy, A., Norvig, P., and Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE intelligent systems*, 24(2):8–12.

Hamdan, M., Rogers, J., and Hamdan, A. (2017). Build to order supply chain efficiency using stochastic frontier analysis (sfa). page 2205 – 2215. Cited by: 0.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition.

Herbrüggen, H. S. (1991). *Modellanalyse von Materialflußsystemen für eine kundennahe Produktion eine empirische Untersuchung*. PhD thesis. ISBN: 3890122345 Dis. an der Uni Passau.

Hilti (2024). Financial report 2023.

Huang, S. H., Uppal, M., and Shi, J. (2002). A product-driven approach to manufacturing supply chain selection. *Supply Chain Management: An International Journal*, 7(4):189–199.

Issah, I., Appiah, O., Appiahene, P., and Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. *Decision Analytics Journal*, 7:100204.

James, G., Witten, D., Hastie, T., Tibshirani, R., et al. (2013). *An introduction to statistical learning*, volume 112. Springer.

Kärkkäinen, H. and Huhtamäki, J. (2023). Performance pricing modeling process develoment by design science reseach.

Katharakis, G., Katharaki, M., and Katostaras, T. (2014). An empirical study of comparing DEA and SFA methods to measure hospital units' efficiency. *International Journal of Operational Research*, 21(3):341.

Khaire, U. M. and Dhanalakshmi, R. (2022). Stability of feature selection algorithm: A review. *Journal of King Saud University - Computer and Information Sciences*, 34(4):1060–1073.

Kiymaz, H., Haque, S., and Choudhury, A. A. (2024). Working capital management and firm performance: A comparative analysis of developed and emerging economies. *Borsa Istanbul Review*, 24(3):634–642.

Klein, C. (2021). Supply Chain Efficiency: A Methodical Comparison between a Value Curve-based Approach and Data Envelopment Analysis.

Kliangkhlao, M., Limsiroratana, S., and Sahoh, B. (2022). The design and development of a causal bayesian networks model for the explanation of agricultural supply chains. *IEEE Access*, 10:86813–86823.

Kohavi, R. and John, G. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97:273–324.

Kotsiantis, S. B., Zaharakis, I., Pintelas, P., et al. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1):3–24.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621.

Larson, P. D. and Rogers, D. S. (1998). Supply chain management: definition, growth and approaches. *Journal of Marketing Theory and Practice*, 6(4):1–5.

Li, S., Ragu-Nathan, B., Ragu-Nathan, T., and Subba Rao, S. (2006). The impact of supply chain management practices on competitive advantage and organizational performance. *Omega*, 34(2):107–124.

Liu, K. P. and Chiu, W. (2021). Supply chain 4.0: the impact of supply chain digitalization and integration on firm performance. *Asian Journal of Business Ethics*, 10(2):371–389.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc.

L'heureux, A., Grolinger, K., Elyamany, H. F., and Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. *Ieee Access*, 5:7776–7797.

Mariappan, M. B., Devi, K., and Venkataraman, Y. (2023). Predicting order processing times in e-pharmacy supply chains during covid pandemic using machine learning—a real-world study. In *Proceedings of International Conference on Data Science and Applications: ICDSA 2022, Volume 2*, pages 175–197. Springer.

Maulud, D. and Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, 1(2):140–147.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., Grosse-Wentrup, M., and Bischl, B. (2020). General pitfalls of model-agnostic interpretation methods for machine learning models. In *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*, pages 39–68. Springer.

Musacchia, G. (2019). Assessment of Optimal Inventory Levels in Supply Chain: Development of an Innovative Model and Implementation in a Real Case.

Nienhaus, J. (2004). Modelling, analysis, and improvement of supply chains - a structured approach.

Nilashi, M., Abumalloh, R. A., Ahmadi, H., Alrizq, M., Abosaq, H., Alghamdi, A., and Mahmood10, S. (2024). Using dematel, clustering, and fuzzy logic for supply chain evaluation of electric vehicles: A scor model. *AIMS Environmental Science*, 11(2):129–156.

Näf, P. (2015). Supply chain efficiency at hilti – logistical efficiency valuation methods in a theoretical context and application of a new approach at hilti.

Parlak, E. and Pescalli, Z. (2023). Improving the Calculation of the Flow Efficiency through Machine Learning.

Polyzotis, N., Zinkevich, M., Roy, S., Breck, E., and Whang, S. (2019). Data validation for machine learning. *Proceedings of machine learning and systems*, 1:334–347.

Porter, M. E. (2001). The value chain and competitive advantage. *Understanding business processes*, pages 50–66.

Putkivaara, L. (2020). Inventory Optimisation By Means of Multivariate Analysis.

Reddy, D. R. and Kameswari, P. (2004). Working capital management practices in pharma industry: A case study of 'cipla limited'. *Management Accountant*, page 638–44.

Rodrigues, T. (2023). Evaluation of value curve-based flow efficiency models.

Rodríguez-Pérez, R. and Bajorath, J. (2020). Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of computer-aided molecular design*, 34(10):1013–1026.

Rogovschi, N., Kitazono, J., Grozavu, N., Omori, T., and Ozawa, S. (2017). t-distributed stochastic neighbor embedding spectral clustering. *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 1628–1632.

Russell, S. J. and Norvig, P. (2016). *Artificial intelligence: a modern approach*. Pearson.

Saaty, R. (1987). The analytic hierarchy process—what it is and how it is used. *Mathematical Modelling*, 9:161–176.

Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., and Müller, K.-R. (2019). *Explainable AI: interpreting, explaining and visualizing deep learning*, volume 11700. Springer Nature.

Sánchez-Maroño, N., Alonso-Betanzos, A., and Tombilla-Sanromán, M. (2007). Filter methods for feature selection–a comparative study. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 178–187. Springer.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, 2(3):160.

Schaffer, C. (1993). Selecting a classification method by cross-validation. *Machine learning*, 13:135–143.

Scheel, H. (2000). *Effizienzmaße der Data Envelopment Analysis*. PhD thesis.

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., and Xu, X. (2017). Dbscan revisited, revisited: why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, 42(3):1–21.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611.

Shapley, L. S. et al. (1953). A value for n-person games.

Shirota, Y., Fujimaki, M., Tsujiura, E., Morita, M., and Machuca, J. A. D. (2021). A shap value-based approach to stock price evaluation of manufacturing companies. In *2021 4th International Conference on Artificial Intelligence for Industries (AI4I)*, pages 75–78. IEEE.

Sieben, B. (2023). PhD thesis.

Sieben, B., Parlak, E., Rodrigues, T., Näf, P., Bley, N., Klein, C., Wolf, B., Witthaut, M., Costa, F., Carvalho, H., Trucco, P., and Bode, C. (2023). Making supply chains more efficient: A novel and robust value curve-based evaluation methods. *under submission*.

Slomka, M. (1989). *Methoden der Schwachstellen- und Ursachenanalyse in logistischen Systemen eine empirische Untersuchung*. PhD thesis.

Soofi, A. A. and Awan, A. (2017). Classification techniques in machine learning: applications and issues. *J. Basic Appl. Sci*, 13(1):459–465.

Stenwig, E., Salvi, G., Rossi, P. S., and Skjærvold, N. K. (2022). Comparative analysis of explainable machine learning prediction models for hospital mortality. *BMC Medical Research Methodology*, 22(1):53.

Stern, C. and Stalk Jr., G. (1998). *Perspectives on Strategy from the BCG*. Boston Consulting Group, Boston, MA.

Stevens, S. (1997). Supply chain operations reference (scor) model. *Overview from supply chain council benchmark study*.

Štrumbelj, E. and Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, 41:647–665.

Tavana, M., Mirzagoltabar, H., Mirhedayatian, S. M., Saen, R. F., and Azadi, M. (2013). A new network epsilon-based dea model for supply chain performance evaluation. *Computers and Industrial Engineering*, 66(2):501 – 513. Cited by: 95.

Thampi, P. (2018). Logistical Efficiency Drivers in a Theoretical Approach and Application within the Hilti Context.

Theng, D. and Bhoyar, K. K. (2024). Feature selection techniques for machine learning: a survey of more than two decades of research. *Knowledge and Information Systems*, 66(3):1575–1637.

Tufte, E. R. (2001). *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT.

Tukey, J. W. et al. (1977). *Exploratory data analysis*, volume 2. Springer.

Ulrich, H. (1981). Die betriebswirtschaftslehre als anwendungsorientierte sozialwissenschaft. *Die Führung des Betriebes, Stuttgart*, 1981:1–25.

Ulrich, H. (1982). Application-oriented science. *The Company*, pages 1–10.

Van Der Maaten, L., Postma, E. O., van den Herik, H. J., et al. (2009). Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10(66-71):13.

VDI (2015). Performance pricing (PP) - fundamentals and application. Technical Standard 2817 1.

Vinzi, V. E., Chin, W. W., Henseler, J., Wang, H., et al. (2010). *Handbook of partial least squares*, volume 201. Springer.

Vitasek, K. (2013). Council of supply chain management professionals.(2013). *Supply Chain Management Terms and Glossary. En Supply Chain Management Terms and Glossary*, page 222.

Wang, Y., Hulstijn, J., and Tan, Y.-H. (2016). Data quality assurance in international supply chains: an application of the value cycle approach to customs reporting. *International Journal of Advanced Logistics*, 5(2):76–85.

Wolf, B. (2020). Wertkurven in der Hilti AG - Anwendungen von Wertkurven im Supply Chain Controlling.

Wong, W. P. and Wong, K. Y. (2007). Supply chain performance measurement system using dea modeling. *Industrial Management and Data Systems*, 107(3):361 – 381. Cited by: 181.

Zachau, T. (1995). *Prozeßgestaltung in industriellen Anlagengeschäften*. PhD thesis. Zugl. Diss., Univ. München, 1994.

Appendix A

Basic Definitions

This section provides definitions of terms and relevant theoretical principles to ensure a clear understanding of the contents focused throughout the dissertation.

A.1 Supply Chain Management

The concept of supply chain management (SCM) was first introduced by Forrester (1958), exploring how SC reacts to fluctuations in demand. The author was able to show that the dynamic complexity involved in moving demand from the customers, through the SC to manufacturers and raw material suppliers distorts the demand patterns. Thus, it was possible to prove the critical interdependence between all the SC stakeholders regarding information flows.

Over the years, a lot of definitions for SCM were developed but it can be agreed that "Supply Chain Management is the coordination of activities, within and between vertically linked firms, for the purpose of serving end customers at a profit" (Larson and Rogers, 1998).

Even before SCM was seen as a critical tool to ensure a competitive advantage, SC managers recognized the value of information and time (Li et al., 2006). It is crucial to reach the markets with the right products faster than the competition, to be able to satisfy customer necessities and to make sure that SC can be synchronized to meet demand fluctuations (Stern and Stalk Jr., 1998).

In the contemporary business environment, SCM plays a pivotal role in integrating key business operations and processes within and across organisations, thereby establishing an efficient business model. It facilitates the coordination of all processes and activities that can, directly or indirectly impact the SC (Vitasek, 2013).

As evidenced by Chopra and Meindl (2006), the strategic management of SC is of paramount importance in tailoring SC to meet specific business needs. By ensuring the strategic alignment between the SC and the business strategy, it is possible to reduce costs and lead times, thereby enhancing the overall performance of the company.

A.2 Supply Chain Efficiency

Prior to addressing the concept of efficiency in the context of SC, it is first necessary to provide a universal definition.

According to Cooper et al. (2003), efficiency refers to the measurement of the output-input ratio within a system. Relative here indicates that an efficiency value must always be evaluated in comparison with the efficiency values of other systems (Dyckhoff, 2003). If a system can generate more output or achieve the same output with less input than other systems, it is deemed efficient (Scheel, 2000).

A.2.1 Product Effects in Supply Chain Efficiency

1 Fisher (1997) was one of the first authors to provide an efficiency-focused framework that supports the selection of the most suitable SC strategy taking into account the conditions of the market under analysis. It splits products into two categories: inventive and utilitarian. According to Fisher (1997), an efficient SC approach is better for functional items, while a responsive SC works best for innovative products. Fisher's framework matrix for linking SC with goods is shown in Figure A.1.

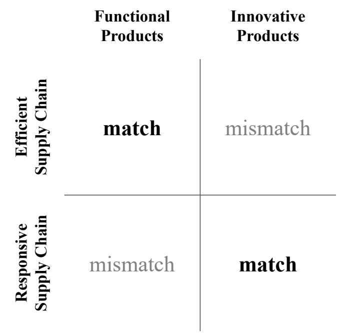


Figure A.1: Fisher (1997) divides products into two categories and defines the supply chain configuration to match them in the most effective way (Fisher, 1997) (modified).

Supported by the work developed by Fisher (1997), Huang et al. (2002) introduced a third product category and adjusted the nomenclature for the SC types. The products can be innovative, hybrid, or functional, and the SC can be lean, agile and hybrid. Figure A.2 shows the matrix with the respective matches between the kind of item and the desired SC configuration type.

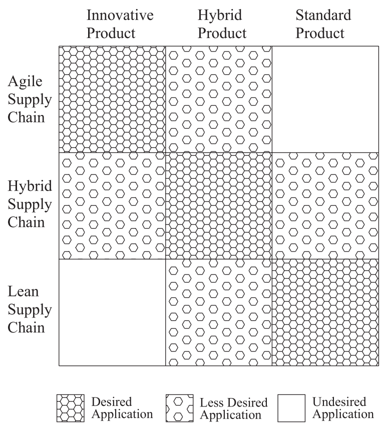


Figure A.2: Huang et al. (2002) extend Fisher (1997) principle to three types of products and supply chain strategies (Huang et al., 2002).

A.2.2 Cost-Responsiveness Efficiency Frontier

Understanding the SC and placing it along the responsiveness spectrum is among the most crucial steps in achieving strategic alignment (Chopra and Meindl, 2006). The cost-responsiveness efficient frontier represents a curve describing the lowest possible cost for a given level of responsiveness. Companies not positioned on this efficient frontier can enhance both their responsiveness and cost performance by moving in that direction. However, a company already situated on the efficient frontier can only improve its responsiveness by increasing costs, which may result in a loss of efficiency (Chopra and Meindl, 2006). The visual depiction of the cost-responsiveness efficiency frontier can be seen in Figure A.3.

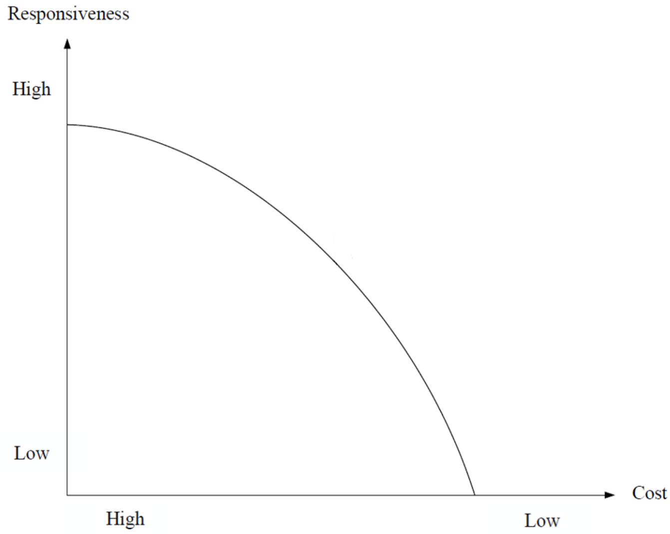


Figure A.3: Cost-responsiveness efficiency frontier shows the trade-off between cost and efficiency both for companies on and off the frontier (Chopra and Meindl, 2006).

Every SC faces a strategic decision regarding its desired level of responsiveness, considering the trade-off between cost and efficiency. While some SC prioritize minimizing costs, others prioritize maximizing responsiveness. Responsive SC are designed to effectively manage both demand and supply fluctuations, whereas efficient SC prioritize cost reduction by sacrificing some flexibility to adapt to varying conditions (Gunasekaran et al., 2008).

A.2.3 Working Capital

Effective working capital management significantly influences the profitability of businesses. Poor management leads to capital being tied up in idle assets, thereby reducing liquidity and profitability (Reddy and Kameswari, 2004). Consequently, it’s crucial for businesses to maintain a working capital level that optimizes costs and benefits while maximizing overall value. Research suggests that as working capital increases up to a certain threshold, corporate performance improves. However, beyond this threshold, the relationship between working capital and performance may become negative (Deloof, 2003).

Achieving an optimal level of working capital involves striking a balance between costs and benefits to maximize firm performance. Managers may initially prefer to increase working capital to boost sales and take advantage of early payment discounts from suppliers. However, excessive investment in working capital beyond the optimal level can lead to increased interest costs, higher bankruptcy risk, and greater credit risk for businesses. Therefore, it's imperative for firm managers to maintain this optimal level and proactively prevent deviations that can negatively impact the firm value (Kiymaz et al., 2024).

A.3 Performance Pricing

Performance Pricing is a statistical approach used to compare purchase prices in the context of supply. The technique is defined in the technical standard published by The German Engineers Association VDI (2015). The goal of performance pricing is to contrast the product price provided by the supplier with an estimated price (or technical value) based on certain value drivers identified by supply teams.

Depending on the product, these value drivers may include characteristics like colour, weight, thickness, size, toughness, material, and others. Based on their impact on the pricing, the projected price will be determined. Illustrations of how one driver and price relate are shown in Figure A.4. Through performance pricing, all value drivers can be compared simultaneously (VDI, 2015).

Figure A.4: Example of plots for 2 drivers and their influence in price (Sieben, 2023).

Value drivers can be categorized as input-oriented or output-oriented, depending on whether they are viewed from the perspective of manufacturers or customers, encompassing both quantitative and qualitative factors. To develop a model, these parameters are inputted into a specialized program that performs the desired computations.

The selection of the most critical value drivers is guided by statistical indicators such as adjusted R^2 and Q^2, depending on the selected algorithm. The program conducts multivariate regression analysis to establish the relationship and weighting between each product and the value drivers. This analysis calculates the impact of each value driver on the price, informing the final equation used in the model generation process (Kärkkäinen and Huhtamäki, 2023).

A value graph is then used as a visual representation of the comparison between the predicted price of a product, and the actual price provided by the supplier, as shown in Figure A.5.

Products, represented by dots, that are perfectly aligned with the 45-degree line have supplier prices matching the projected price, indicating a relatively appropriate price level. Products above the line are more expensive than anticipated, while those below the line offer better value to the customer. This suggests customers receive a valuable product for less money than statistically predicted. Conducting this research allows buyer-side supply teams to make informed decisions and engage in discussions with suppliers confidently.

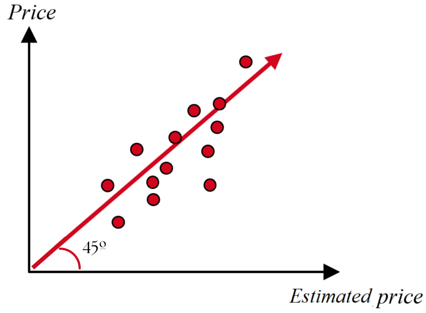


Figure A.5: Example price versus estimated price (Sieben, 2023).

A.4 Value Curves

Widely used in industry, value curves are an easily comprehensible graphical method for estimating the appropriateness of a production or logistics setup. Value curves use the Bill of Materials (BoM) as a basis. They depict the throughput time of the respective production and logistics steps on the abscissa and the according cumulated cost of each production step, i.e., manufacturing, assembly, transportation, or storage, on the ordinate, as described in Figure A.6

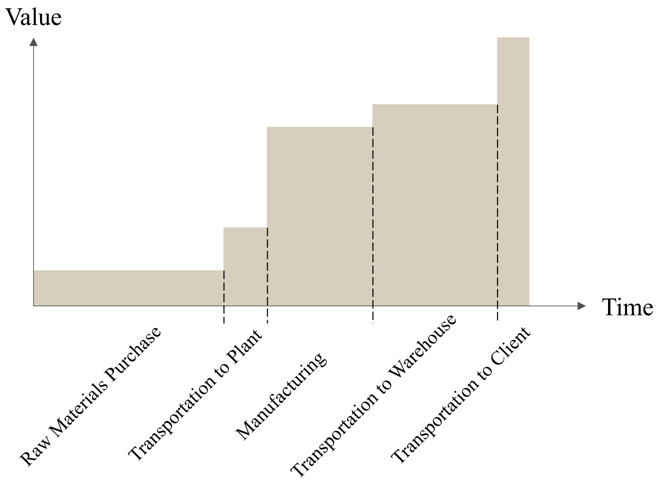


Figure A.6: Example of a 5-step value curve representing different stages of the supply chain (Rodrigues, 2023).

Cost curves are specific to the combination of product items and production lines. For that reason, we refer to item-line combinations for the particular production setup for an item. Even if two different products were to be produced on the same production line, still two item-line combinations and therefore two value curves would result. The area covered by all sequential steps is proportional to the capital tied up in the production chain for a particular product: An efficient item-line combination would thus reach the end-product value while covering an area as small as possible (Danert, 1988; Nienhaus, 2004).

Within the SC, value curve analysis aids in strategic decision-making and process improvement. Businesses can make informed decisions regarding investments in SC activities while identifying collaboration opportunities (Porter, 2001).

Despite widespread industry adoption since the late 1980s (Zachau, 1995), literature on value curves remains relatively scarce (Sieben, 2023). Several doctoral theses have explored this topic qualitatively (Slomka, 1989; Herbrüggen, 1991; Zachau, 1995). However, no quantification methods were further provided.

A.5 Flow Efficiency Calculation

The term "flow efficiency" was introduced by Sieben (2023), representing the calculated efficiency of a value curve. Nienhaus (2004) initially highlighted the quantitative use of the area occupied by the value curve graph in relation to the total area. Subsequently, within Sieben's research project, various weighting approaches for flow efficiency calculation were proposed, with contributions from several authors. (Näf, 2015; Thampi, 2018; Wolf, 2020; Gomes, 2021; Bley, 2022; Parlak and Pescalli, 2023; Rodrigues, 2023). These contributions include the development of new calculation methods (Näf, 2015; Sieben, 2023; Rodrigues, 2023) and the application to diverse product portfolios (Thampi, 2018; Gomes, 2021; Bley, 2022; Rodrigues, 2023). This section provides an overview of some of these significant contributions.

A.5.1 Nienhaus' efficiency model

In a sidenote of his doctoral dissertation, Nienhaus (2004) suggested that efficiency, referred to here as flow efficiency, could be deduced from value curves. According to him, the flow efficiency of a SC, depicted by a value curve, could be calculated as the difference between one and the quotient of the area occupied by the value curve graph and the total area, as illustrated in Figure A.7 and Equation A.1.

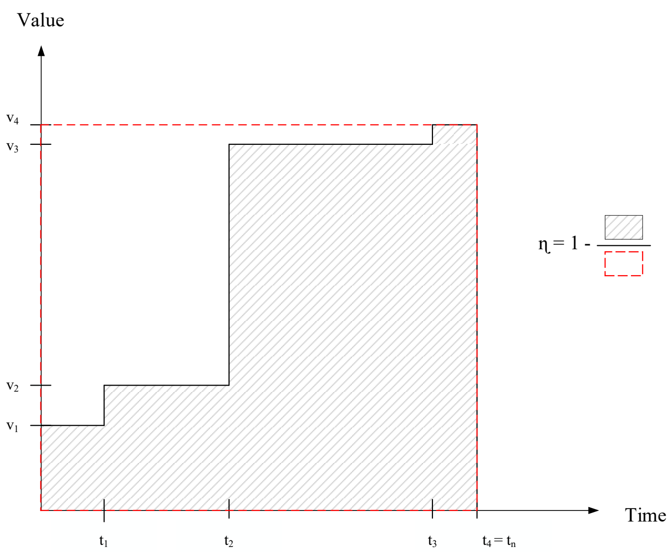


Figure A.7: Graphical representation of Nienhaus' model (Sieben, 2023; Rodrigues, 2023).

$$\eta = 1 - \frac{\sum_{i=1}^{n} v_i \times (t_i - t_{i-1})}{t_n \times v_n} \tag{A.1}$$

The underlying principle of this approach is that a lower filled area on the graph indicates lower capital investment and therefore higher efficiency. However, as noted by Näf (2015), this assumption does not hold true in all scenarios. For instance, when the initial step of the value curve involves a longer duration and lower value compared to subsequent steps, efficiency may be misrepresented and consequently misinterpreted.

A.5.2 Näf's Model

Acknowledging the limitation of Nienhaus' model, Näf (2015) developed an alternative method aimed at identifying scenarios where such limitations occur. In such cases, the proposed method suggests disregarding the initial step in the value chain efficiency calculation. This heuristic approach suggests that if the duration of the first step of the curve is four times longer than the subsequent steps, then the first step is excluded from the calculation, and the Nienhaus method is applied only to the remaining steps. This method is described by Equation A.2.

$$\eta = 1 - \frac{\sum_{i=1}^{n} v_i \times (t_i - t_{i-1})}{t_n \times v_n} \quad with \quad \frac{t_i}{t} \geq z \times \frac{v_i}{v} \quad and \quad z = 4 \tag{A.2}$$

A.5.3 Sieben's Models

Sieben (2023) introduced alternative approaches, including the calculation of partial flow efficiency for individual steps of the value curve. To assess the partial flow efficiency, the author begins with the assumption that a Dirac-like bump function in the value curve represents the ideal production setup from a flow efficiency perspective, as all value is added immediately without binding any capital.

An efficient step, defined as one in which a substantial amount of capital is invested in a relatively short period of time, may have the opposite effect to that intended for an efficient process in the initial phase of the SC. Furthermore, the efficiency of a step that has a significantly longer duration than all the others may have a much larger impact on total flow efficiency than the remaining steps. Recognizing these nuances and the positive impact of high flow efficiency in later stages of the SC, Sieben (2023) developed models employing weighting approaches. These methods mathematically express the relative importance of each step in the value curve. By accounting for these considerations, the models aim to provide a more accurate representation of SC efficiency dynamics, ensuring that later stages receive appropriate emphasis in the analysis.

Sieben (2023) proposed eight different weighting models: value-based, time-based, combined value-time-based, value vector-based, horizontal slack-based, vertical slack-based, exponential and gravitational distance.

The existing proposals for flow efficiency calculation, lacked thorough testing with sufficient case data to assess their validity. To bridge this gap, Rodrigues (2023) collected item production data for more than 400 cases and applied an established evaluation framework for the first time. The results showed that vertical slack-based weighting performed the best for the cases under analysis.

A.5.3.1 Vertical Slack-based Weighting

In this case, the vertically downstream areas originated by a production step when determining the weighting are considered. With this approach, a high partial flow efficiency directly contributes to a positive effect on the underlying surface, therefore the respective vertical residual surface should be positively related to the total vertical residual surface. This is visually represented in Figure A.8 and translated by Equation A.3.

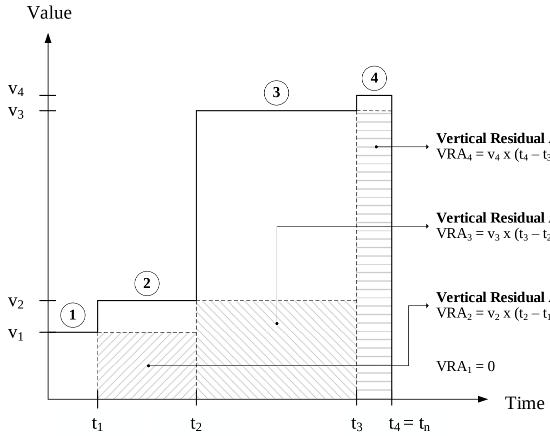


Figure A.8: Representation of vertical slack-based weighting (Sieben, 2023; Rodrigues, 2023).

$$\eta = \sum_{i=1}^{n} \left\{ arctan\left(\frac{t_n \cdot (v_i - v_{i-1})}{v_n \cdot (t_i - t_{i-1})}\right) \cdot \frac{2 \cdot v_{i-1} \cdot (t_i - t_{i-1})}{\pi \cdot \sum_{i=1}^{n} (v_{i-1} \cdot (t_i - t_{i-1}))} \right\} \tag{A.3}$$

A.6 Statistical Tests

In the realm of statistical analysis, determining the distribution and nature of your data is a fundamental step that influences the choice of subsequent tests for analyzing significant differences (Fisher, 2006). This section delves into the essential concepts for assessing data normality and testing for significant differences between datasets.

A.6.1 Data Normality Assessment

Understanding the distribution of your data is crucial for selecting appropriate statistical tests (Fisher, 2006). A dataset is considered normally distributed if it follows a bell-shaped curve known as the Gaussian distribution or normal distribution (Fisher, 2006). This distribution is characterized by several key properties:

- Symmetry: The distribution is symmetric around the mean, meaning the left and right sides of the curve are mirror images.
- Central Tendency: The mean, median, and mode of the distribution are all equal and located at the center.
- Standard Deviation: Approximately 68% of the data falls within one standard deviation of the mean, about 95% within two standard deviations, and nearly 99.7% within three standard deviations.

The normal distribution is crucial in statistics due to the Central Limit Theorem (CLT) (Fischer and Fischer, 2011). The CLT states that, under certain conditions, the sum or average of a large number of independent, identically distributed random variables will be approximately normally distributed, regardless of the underlying distribution. This theorem justifies the use of normal distribution assumptions in many statistical tests and methods, as it allows for the application of these methods to a wide range of problems (Fischer and Fischer, 2011).

The Shapiro-Wilk test is a widely used method for evaluating whether a sample comes from a normally distributed population(Shapiro and Wilk, 1965). This test provides a p-value:

- A p-value greater than 0.05 suggests that the data does not significantly deviate from a normal distribution.
- A p-value less than 0.05 indicates that the data significantly deviates from a normal distribution, suggesting non-normality.

The Shapiro-Wilk test is particularly powerful for small sample sizes, making it a preferred choice in many research scenarios (Shapiro and Wilk, 1965).

A.6.2 Statistical Tests for Significant Differences

Once the distribution of the data is established, researchers can apply appropriate statistical tests to determine significant differences between groups. The choice of test depends on whether the data meets the assumption of normality.

ANOVA (Analysis of Variance): For normally distributed data, ANOVA is the go-to method for comparing the means of three or more independent groups (Fisher, 2006). It assesses whether any observed differences in means are statistically significant or if they could have occurred by random chance. ANOVA helps in determining if at least one group mean is different from the others, providing insight into potential factors influencing the data.

Kruskal-Wallis Test: When the data does not follow a normal distribution, the Kruskal-Wallis test serves as a robust non-parametric alternative to ANOVA (Kruskal and Wallis, 1952), also comparing the medians of three or more independent groups. Unlike ANOVA, the Kruskal-Wallis test does not assume normality and is less affected by outliers and heteroscedasticity (unequal variances). It ranks the data and evaluates whether the distribution of ranks differs significantly between groups.

A.7 Artificial Intelligence and Machine Learning

In today's data-driven world, every action generates vast amounts of data. Artificial Intelligence (AI), a field of computer science, encompasses systems that mimic human-like functions and continuously learn from their actions to improve performance (García-Arca et al., 2016). Machine Learning (ML), a subset of AI, is particularly prominent due to its focus on analysing data and deriving insights without explicit programming (Alzubi et al., 2018). This ability is transforming various sectors, with businesses leveraging AI technologies to streamline operations, personalize customer experiences, and drive innovation. In industrial settings, AI technologies play a crucial role in sensing, analyzing, and interpreting data to solve complex problems, while uncovering opportunities to optimize processes(Bharadiya, 2023).

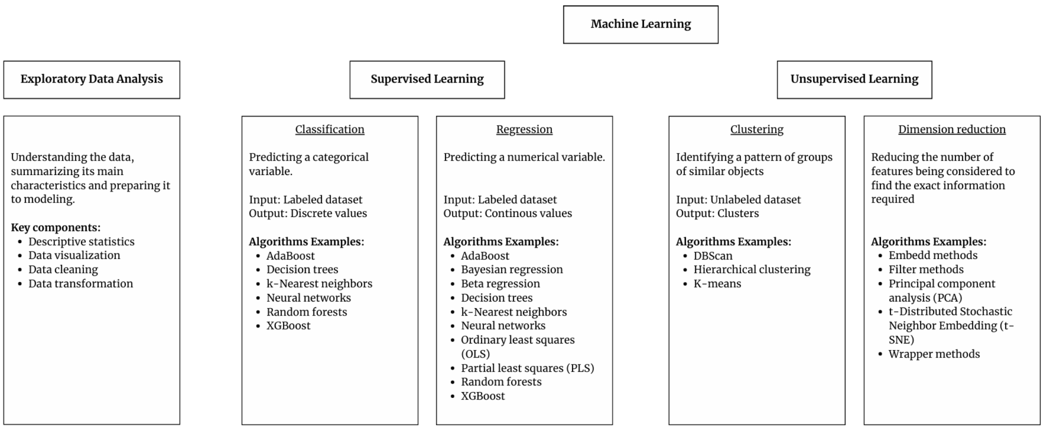


Figure A.9: Classification of machine learning algorithms and exploratory data analysis(The author).

This section will first delve into Exploratory Data Analysis (EDA), describing its importance for data pattern discovery. Secondly, it focuses on ML, highlighting the key characteristics of the learning methods. When discussing ML, three main categories of learning techniques emerge: supervised learning, unsupervised learning, and reinforcement learning (Sarker, 2021). Figure A.9 illustrates the specific application areas of two of these three techniques while providing examples. Considering the scope of the methodological development presented in the following sections, supervised and unsupervised learning models appear to be the most suitable choices for achieving the final objectives of this dissertation.

A.7.1 Exploratory Data Analysis

EDA can be described as a combination of statistical analysis and data visualisation techniques to develop a better understanding of the key characteristics, patterns and potential issues in the data (Tukey et al., 1977). EDA is normally used to reveal hidden patterns and relationships, identify errors and inconsistencies, formulate and test hypotheses and, ultimately, prepare the data for advanced analysis (Tukey et al., 1977). By using EDA, it is possible to gain a deeper understanding of the data while enabling informed decision-making (Brehmer, 1992).

Within this field, it is worth highlighting some techniques used for multivariate analysis. These techniques are the ones that better suit this thesis's purpose as they enable the simultaneous visualization of all numerical relationships across the full data set.

- **Heat Map Matrix:** visual representation of multivariate data organized as a matrix of rows and columns, with each cell colour-coded to reflect the degree of correlation between variables. By converting the correlation matrix into a colour gradient, the heat map facilitates the identification of patterns and relationships among variables. This visualization aids in the identification of the most relevant attributes for constructing accurate machine learning models (Tufte, 2001).

- **Scatter Plot Matrix:** graphical representation that displays the relationships between multiple pairs of variables in a dataset. It consists of a grid of scatter plots, where each scatter plot represents the relationship between two variables. The variables are typically plotted on a two-dimensional coordinate system, with one variable on the x-axis and another on the y-axis. This technique is highly effective for visually examining the relationships or trends between variables in a dataset (Tufte, 2001).

A.7.2 Supervised Learning

Supervised learning relies on labelled datasets to guide algorithms in classifying data or predicting outcomes when new inputs are introduced (Kotsiantis et al., 2007). These labelled inputs and corresponding outputs enable the model to continuously refine its accuracy over time. Supervised learning can be categorized into classification and regression techniques, each serving distinct purposes:

Classification: This technique involves categorizing test data into predefined groups using algorithms. Common classification methods include decision trees, support vector machines, random forests, and linear classifiers (Soofi and Awan, 2017).

Regression: In regression, algorithms are utilized to understand the relationship between independent and dependent variables, making predictions with numerical datasets. This is particularly useful when forecasting numerical outcomes, such as predicting the efficiency of a SC. Polynomial regression, logistic regression, and neural network regression models are among the commonly employed regression algorithms (Maulud and Abdulazeez, 2020).

Throughout the following sections, machine learning concepts and techniques, crucial to this dissertation's methodological development, are described. This will provide the reader with the necessary foundational understanding.

A.7.3 Unsupervised Learning

Unsupervised learning techniques encompass a set of algorithms used to analyze and extract patterns from unlabeled data. These methods uncover hidden relationships within the dataset without the need for target variables. Instead, they rely on intrinsic data properties to cluster similar data points together. Unsupervised learning techniques play a crucial role in tasks such as clustering and dimensionality reduction (Barlow, 1989).

Clustering

Unlabelled data can be grouped using the data mining approach of clustering, based on some thresholds that identify the similarities or differences of the clusters (Abonyi and Feil, 2007). DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a clustering algorithm that identifies clusters based on the density of data points. Data points are categorised as core, boundary or noise points (Schubert et al., 2017). Core points are those with a sufficient number of neighbours within a specified radius, while boundary points are those within the radius of a core point but without sufficient neighbours. Noise points do not belong to any cluster (Schubert et al., 2017). DBSCAN is effective for clustering datasets with different shapes and densities, as well as in the presence of noise or outliers (Ertöz et al., 2003).

A.7.4 High Dimensionality

High dimensionality in a dataset poses significant challenges to ML models' ability to accurately predict outcomes. When there are too many attributes, it becomes harder to discern meaningful patterns from noise (L'heureux et al., 2017). Dimensionality reduction techniques address these challenges by transforming the high-dimensional data into a lower-dimensional space while preserving essential information (Ayesha et al., 2020). By eliminating redundant dimensions and retaining only the most informative ones, dimensionality reduction methods simplify the model's representation of the data, making it easier to learn and generalize from (Van Der Maaten et al., 2009). This process not only improves prediction accuracy but also enhances computational efficiency and ensures better interpretability of the results (Van Der Maaten et al., 2009).

Feature selection and dimensionality reduction are two effective strategies for mitigating the challenges posed by high dimensionality in machine learning (Bolón-Canedo et al., 2016).

Feature Selection: involves identifying and selecting the most relevant explanatory variables from the original dataset while discarding irrelevant or redundant ones (Cai et al., 2018). There are three categories of feature selection techniques:

- Filter Methods: evaluate the relevance of the features independently of the learning algorithm (Sánchez-Maroño et al., 2007). Person correlation coefficient, information gain and R^2 test are among the techniques of this group (Chandrashekar and Sahin, 2014).
- Wrapper Methods: iterates through various feature combinations to identify the optimal subset (Kohavi and John, 1997). Forward selection, backward elimination and recursive feature elimination are some of the examples of the techniques in this category (Chandrashekar and Sahin, 2014).
  - Forward stepwise selection: starts with an empty model, adding one feature at a time that improves the model the most until no further significant improvement is possible. (Kohavi and John, 1997)
  - Backward stepwise selection: starts with the full model, removing one feature at a time that contributes the least until further removal would degrade the model's performance. (Kohavi and John, 1997)
- Embedded Methods: integrate feature selection directly into the model training process. LASSO and ridge regressions are the most commonly used (Chandrashekar and Sahin, 2014).

Dimensionality Reduction: involves transforming the original high-dimensional dataset into a lower-dimensional representation while preserving essential information (Van Der Maaten et al., 2009).

The following techniques are worth to highlight:

- Principal Component Analysis (PCA): transforms the original variables into a set of linearly uncorrelated variables called principal components (Abdi and Williams, 2010). PCA begins by computing the correlation matrix of the data set and then calculates the eigenvectors and eigenvalues of that matrix. Then it selects the eigenvectors with the largest eigenvalues as the principal components and projects the original data onto these components (Abdi and Williams, 2010).

- t-Distributed Stochastic Neighbor Embedding (t-SNE): projects high-dimensional data onto a lower-dimensional space (usually 2 or 3). It tries to preserve the distances between similar points in the high-dimensional space while also separating dissimilar points in the lower dimension (Rogovschi et al., 2017).

A.8 Data Validation

Once a ML model is trained, its performance on new, unseen data cannot be guaranteed (Issah et al., 2023). Therefore, it is uncertain whether the model will maintain the expected accuracy and variance in a production scenario (Issah et al., 2023). To ensure the reliability of the forecasts generated by the ML model, it must undergo a validation process which involves determining whether the model's numerical results, quantifying relationships between variables, are accurately reflecting the data (Polyzotis et al., 2019).

To assess a model's performance and determine if it is well-generalized or suffering from underfitting or overfitting, it must be tested on data that was not used during training (Dietterich, 1995). Cross-validation (CV) is a key method for evaluating ML models, especially useful when data is limited. In CV, a portion of the data is set aside for testing and validation, separate from the training data. This process is repeated iteratively to achieve accurate training and testing operations (Schaffer, 1993). Below, we will discuss two of the most common validation approaches.

A.8.1 Test and Train Datasets

Instead of employing complex validation methods, a simpler and less computationally intensive approach for validation is the Train/Test split (Russell and Norvig, 2016). In this method, the complete dataset is randomly divided into a training set and a test set, typically in proportions of 70:30 or 80:20. The training set is used to train the model, while the test set is used for validation (Hastie et al., 2009). However, this straightforward approach carries a significant risk of bias, particularly when dealing with limited data, as the rows might vary extensively and important information might be excluded from the training process (Molnar et al., 2020). If the dataset is sufficiently large and both the training and test sets have similar distributions, this method can be considered valid (Hastie et al., 2009).

A.8.2 K-Folds Cross Validation

The K-Folds cross-validation technique typically produces a model with less bias compared to simpler data splitting methods (Molnar et al., 2020). Due to its iterative nature, it ensures that every observation from the original dataset has a chance of appearing in both the training and test sets (Hastie et al., 2009). This method is particularly useful when the available data is limited, as it helps mitigate overfitting by resampling the data. According to Hastie et al. (2009), the steps for K-Folds cross-validation are as follows:

- Splitting the Data: The full dataset is randomly divided into K folds, where K typically ranges from 5 to 10 depending on the data size. A larger value of K generally reduces bias but makes the approach more similar to a simple train-test split as K decreases.
- Training and Validation: K-1 folds are used for training, while the remaining Kth fold is used for validation. The error (E) for each iteration is recorded.
- Iteration and Averaging: The process is repeated until each of the K folds has been used as the test set. After all iterations, the average of the recorded error values is calculated to serve as the model's performance metric, as shown below:

$$E = \frac{1}{K}\sum_{i=1}^{K} E_i \tag{A.4}$$

This average error provides a comprehensive measure of the model's performance across all folds (Hastie et al., 2009).

A.9 Model's Evaluation

Evaluating the performance of ML models is a critical aspect of the model development process, providing information on the model's capacity to make accurate predictions (Bishop, 2006). In the context of regression problems, various metrics are employed to assess the model's performance, each offering different insights into the model's predictive capabilities:

- **Coefficient of determination - R^2:** measures the proportion of the variance in the dependent variable that is predictable from the independent variables (James et al., 2013). Particularly useful for understanding the proportion of variance explained by the model, which is crucial for models where the goal is to capture as much of the variability in the data as possible (Bishop, 2006). It provides an indication of how well the model fits the data, with values ranging from 0 to 1. A higher R^2 value indicates a better fit. The formula for R^2 is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{A.5}$$

- **Mean Absolute Error - MAE:** measures the average magnitude of the errors in a set of predictions, without considering their direction. It provides a straightforward interpretation of the prediction errors in the same units as the output variable (James et al., 2013). The formula for MAE is:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{A.6}$$

- **Mean Squared Error - MSE:** calculates the average of the squares of the errors, giving more weight to larger errors. This makes MSE particularly useful when we want to penalize large errors (James et al., 2013). The formula for MSE is:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{A.7}$$

- **Root Mean Squared Error - RMSE:** derived from MSE, it provides an error metric that is in the same units as the output variable. By taking the square root of the MSE, RMSE offers an interpretable metric that can be directly compared to the actual values (James et al., 2013). The formula for RMSE is:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{A.8}$$

Evaluating these metrics collectively can provide a comprehensive view of the model's strengths and weaknesses. For instance, a model with a high R^2 but also a high RMSE might be capturing the overall trend well but struggling with specific outliers or large deviations. Conversely, a low MAE but a high R^2 might indicate that while the model is generally accurate, it fails to explain much of the variability in the data (Bishop, 2006).

A.10 Machine Leaning Models' Interpretability

Interpreting the outcomes of machine learning models often poses a challenge, particularly when attempting to derive actionable conclusions from the predictions, underscoring the significance of model interpretability. Visualizing the importance of explanatory variables in ML predictions enhances interpretability, thereby facilitating the extraction of insights that support informed decision-making (Molnar, 2020). Moreover, such visualizations support feature selection by identifying the most relevant variables and excluding those that minimally contribute to the model's performance, enhancing performance and reducing complexity (Samek et al., 2019). Additionally, it reveals underlying data patterns, offering valuable information for model refinement and the potential discovery of domain-specific knowledge (Hastie et al., 2009).

SHAP (SHapley Additive exPlanations) values offer a robust and theoretically grounded approach to interpreting ML model predictions, being one of the most commonly used techniques in this context (Rodríguez-Pérez and Bajorath, 2020). Derived from cooperative game theory, SHAP values provide a method for fairly distributing the total gains (or payout) among players based on their contributions to the overall success of the coalition (Molnar, 2020). In the context of ML, these "players" are the features of the model, and the "payout" is the prediction made by the model (Štrumbelj and Kononenko, 2014).

The Shapley value for a feature is the average of its marginal contributions across all possible subsets of features. This involves calculating how the prediction changes when the feature is added to subsets of other features, reflecting its contribution to the prediction (Shapley et al., 1953). The formal Shapley value formula is presented in Equation A.9.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left[v(S \cup \{i\}) - v(S) \right] \tag{A.9}$$

where:

- ϕ_i is the Shapley value for feature i.
- N is the set of all features.
- S is a subset of N not containing i.
- $v(S)$ is the value (prediction) for the subset S.

According to Štrumbelj and Kononenko (2014), Shapley values' properties offer a rigorous and versatile approach to interpreting machine learning models:

- **Efficiency:** The sum of the Shapley values for all features equals the total value of the game, which in the context of ML is the model's prediction. This property ensures that the contributions of all features are fairly distributed and collectively account for the entire prediction. This is crucial for model interpretability as it guarantees that all factors influencing the prediction are considered and quantified, providing a comprehensive understanding of how each feature contributes to the model's output.
- **Symmetry:** If two features contribute equally to all coalitions, they receive equal Shapley values. This property ensures that the interpretability method is fair and unbiased, treating features with equivalent impact equally.
- **Dummy:** Features that do not change the prediction when added to any subset receive a Shapley value of zero. This property ensures that irrelevant features are appropriately identified and given no weight in the interpretability analysis. This means that features which do not contribute to the model's predictions are recognized and can be safely ignored, simplifying the model.
- **Additivity:** The Shapley value for combined games (or predictions from combined models) is the sum of the Shapley values from the individual games. This property allows Shapley values to be naturally extended to ensemble models, which are commonly used in machine learning. For example, in a random forest or gradient-boosting model, the overall feature importance can be derived by summing the Shapley values from each individual tree, providing a consistent measure of feature contributions across complex models.

It is essential to take into consideration that implementing SHAP in practice can be computationally intensive due to the need to calculate the contribution of each feature across all possible subsets of features, which grows exponentially with the number of features (Lundberg and Lee, 2017). This computational complexity implies the use of approximation techniques to make SHAP feasible for large datasets and complex models. The Python package "shap" provides efficient algorithms to approximate Shapley values, significantly reducing computation time. However, these approximations have drawbacks, particularly the assumption that features are independent. In datasets with high multicollinearity, where features are highly correlated, this assumption can lead to unreliable results as the interdependencies between features are not accurately captured (Lundberg and Lee, 2017). Consequently, while "shap" makes the implementation of Shapley values practical, caution is needed when interpreting results from datasets with significant feature dependencies. Lundberg and Lee (2017) proposes the analysis of SHAP interaction values to understand interdependencies between features and mitigate the risks associated with multicollinearity as these values measure how the impact of one feature on the prediction changes depending on the value of another feature (Lundberg and Lee, 2017).

The beeswarm plot serves as a widely employed visualization method to illustrate SHAP values and facilitate the interpretation of ML model predictions. By plotting the SHAP values of each feature for every sample, this visualization offers a comprehensive overview of the importance of features in the model's decision-making process. In Figure A.10, features are sorted based on the sum of SHAP value magnitudes across all samples, allowing for a clear depiction of each feature's impact on the model output. Additionally, the colour gradient in the plot represents the feature values, with red indicating high values and blue indicating low values. This colour scheme provides further insight into the relationship between feature values and their corresponding SHAP values.

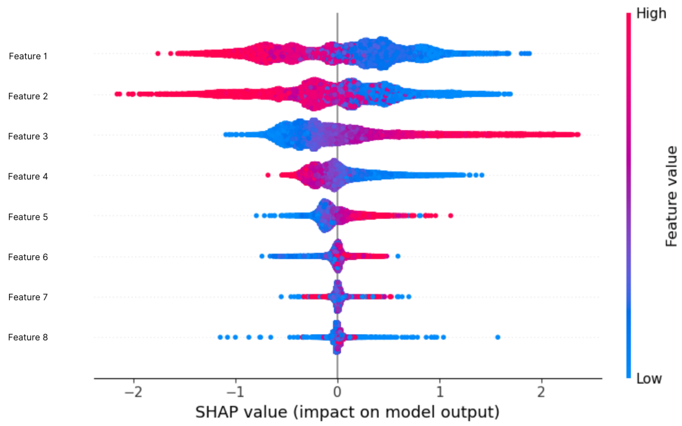


Figure A.10: Beeswarm plot for Shap Values visualization (The author).

Appendix B

Data Availability Evaluation Matrix Applied to the Selected Flow Efficiency Drivers

			Purchase Price	Selling Price	Weight	Volume	PVD	Unit Production Cost	Transportation Costs	Turnover	Delivery Reliability	Annual demand	Throughput	Throughput variance	Lead Time	BI Risk
If data is available	Accuracy of the data	Semantic	High	High	High	High	High	Medium	-	-	High	High	High	High	High	-
		Syntactically	High	High	High	High	High	High	-	-	High	High	High	High	High	-
	Completeness of the data	Missing data	Medium	Medium	Medium	Medium	Medium	Low	-	-	High	Low	High	High	Medium	-
		Duplicates	Low	Low	Low	Low	Low	Low	-	-	Medium	Medium	Low	Low	Low	-
	Objectivity of the data		Low	Medium	High	High	High	High	-	-	Medium	High	High	High	High	-
If data is missing	Effort of the survey		-	-	-	-	-	-	High	High	-	-	-	-	-	High
	Relevance of the efficiency driver		-	-	-	-	-	-	High	High	-	-	-	-	-	Low
VDI Rating			Optional	Optional	Important	Important	Important	Important	Negligible	Negligible	Negligbe	Important	Negligible	Negligbe	Important	Negligible
Data Suitability			No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Overall Rating			Negligible	Negligible	Important	Important	Important	Important	Negligible	Negligible	Negligible	Important	Negligible	Negligbe	Important	Negligible

			Inventory	MOQ	Batch size	Forecast error	Flexibility	Response Time	Product Commonality	Product Complexity	CODP	Complexity of the supplier network	Working capital	Obsolescence risk	Product lifecycle	Technological Integration
If data is available	Accuracy of the data	Semantic	High	High	High	High	-	High	-	High	-	-	High	-	-	High
		Syntactically	High	High	High	Medium	-	High	-	High	-	-	High	-	-	High
	Completeness of the data	Missing data	Medium	Low	Low	Medium	-	Medium	-	Medium	-	-	High	-	-	High
		Duplicates	Medium	Low	Low	Medium	-	Low	-	Medium	-	-	Medium	-	-	Low
	Objectivity of the data		Medium	High	High	Medium	-	Medium	-	Medium	-	-	Low	-	-	Low
If data is missing	Effort of the survey		-	-	-	-	High	-	High	-	High	High	-	High	High	-
	Relevance of the efficiency driver		-	-	-	-	High	-	High	-	Medium	Medium	-	Medium	High	-
VDI Rating			Important	Important	Important	Optional	Negligible	Important	Negligible	Optional	Negligible	Negligible	Negligible	Negligible	Negligible	Important
Data Suitability			Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No
Overall Rating			Important	Important	Important	Important	Negligible	Important	Negligible	Important	Negligible	Negligible	Negligible	Negligible	Negligible	Negligible

Appendix C

Flow Efficiency Drivers under Analysis

Table C.1: Description of the flow efficiency drivers used in methodological development.

Flow efficiency driver	**Definition**	**Units**
Annual demand	The amount of goods sold each year	Units
Batch size	The number of products produced continuously without pausing	Units
Forecast error	The deviation of forecast demand from actual demand	Percentage (%)
Lead time	The time from order placement to complete delivery	Days
Machine time	Actual time in which the product is processed	Hours
Material cost	The cost of raw materials required to produce a product	Currency
Minimum order quantity (MOQ)	The minimum quantity that can be ordered of a product	Units
Number of production steps	The total number of steps involved in the production process of a product	Steps
Product complexity	Determined by the number of components listed in a product's Bill of Materials (BoM).	Number of components
Production cost	The expenditure incurred in the production of a single unit of output.	Currency
Product value density (PVD)	The value of a product in relation to its weight and volume.	Currency per unit weight or volume
Safety stock	Additional inventory held as a buffer against unexpected fluctuations in demand or supply	Units
Volume	The total volume of a product	Cubic meters (m^3)
Weight	The total weight of a product	Kilograms (kg)

Appendix D

Models' Performance throughout Backward Stepwise Elimination

Table D.1: Models' performance throughout backward stepwise elimination.

Model	# Folds	Metric	It.1	It.2	It.3	It.4	It.5	It.6	It.7	It.8	It.9	It.10	It.11	It.12	It.13	It.14
AdaBoost	20	R2	0.669	0.661	0.658	0.664	0.660	0.662	0.660	0.659	0.653	0.638	0.533	0.464	0.440	0.313
AdaBoost	20	RMSE	0.069	0.070	0.071	0.070	0.070	0.070	0.070	0.071	0.071	0.072	0.082	0.088	0.091	0.101
AdaBoost	10	R2	0.674	0.669	0.670	0.677	0.674	0.680	0.660	0.657	0.659	0.653	0.494	0.467	0.432	0.369
AdaBoost	10	RMSE	0.070	0.070	0.070	0.070	0.070	0.069	0.071	0.071	0.071	0.071	0.087	0.089	0.092	0.096
AdaBoost	5	R2	0.684	0.678	0.675	0.683	0.681	0.680	0.671	0.598	0.654	0.670	0.623	0.395	0.459	0.383
AdaBoost	5	RMSE	0.069	0.070	0.071	0.070	0.070	0.070	0.071	0.078	0.073	0.071	0.076	0.097	0.091	0.097
AdaBoost	3	R2	0.674	0.692	0.689	0.691	0.694	0.674	0.655	0.657	0.650	0.548	0.545	0.388	0.455	0.272
AdaBoost	3	RMSE	0.071	0.069	0.070	0.070	0.069	0.071	0.073	0.073	0.074	0.084	0.084	0.098	0.092	0.107
Bayesian Reg	20	R2	0.229	0.230	0.194	0.194	0.195	0.187	0.180	0.179	0.170	0.169	0.129	0.068	0.051	0.025
Bayesian Reg	20	RMSE	0.107	0.107	0.109	0.109	0.109	0.110	0.110	0.110	0.111	0.111	0.114	0.118	0.119	0.121
Bayesian Reg	10	R2	0.232	0.234	0.204	0.203	0.204	0.195	0.188	0.187	0.178	0.177	0.133	0.069	0.053	0.027
Bayesian Reg	10	RMSE	0.108	0.107	0.110	0.110	0.110	0.111	0.111	0.111	0.112	0.112	0.115	0.119	0.120	0.122
Bayesian Reg	5	R2	0.243	0.245	0.210	0.211	0.211	0.203	0.195	0.195	0.187	0.184	0.133	0.076	0.059	0.032
Bayesian Reg	5	RMSE	0.108	0.108	0.111	0.111	0.111	0.111	0.112	0.112	0.112	0.112	0.116	0.120	0.121	0.123
Bayesian Reg	3	R2	0.249	0.247	0.211	0.212	0.213	0.203	0.199	0.198	0.192	0.188	0.129	0.063	0.046	0.035
Bayesian Reg	3	RMSE	0.109	0.109	0.111	0.111	0.111	0.112	0.112	0.112	0.113	0.113	0.117	0.121	0.122	0.123
Beta Reg	20	R2	0.378	0.375	0.389	0.376	0.367	0.371	0.369	0.312	0.287	0.265	0.254	0.243	0.201	0.134
Beta Reg	20	RMSE	0.1065	0.1066	0.101	0.114	0.121	0.119	0.123	0.123	0.12	0.125	0.126	0.124	0.125	0.124
Beta Reg	10	R2	0.396	0.385	0.383	0.39	0.379	0.365	0.363	0.355	0.323	0.281	0.263	0.251	0.243	0.212
Beta Reg	10	RMSE	0.106	0.107	0.103	0.109	0.111	0.114	0.115	0.117	0.119	0.12	0.122	0.124	0.125	0.128
Beta Reg	5	R2	0.392	0.383	0.38	0.388	0.375	0.365	0.355	0.345	0.295	0.275	0.255	0.245	0.235	0.195
Beta Reg	5	RMSE	0.107	0.108	0.104	0.112	0.113	0.112	0.116	0.118	0.12	0.121	0.123	0.125	0.126	0.129
Beta Reg	3	R2	0.385	0.381	0.375	0.385	0.373	0.355	0.352	0.344	0.293	0.273	0.253	0.242	0.234	0.192
Beta Reg	3	RMSE	0.108	0.109	0.105	0.111	0.113	0.116	0.117	0.119	0.121	0.122	0.124	0.126	0.127	0.134
Decision Trees	20	R2	0.803	0.800	0.833	0.832	0.832	0.831	0.808	0.814	0.803	0.782	0.674	0.610	0.589	0.358
Decision Trees	20	RMSE	0.053	0.054	0.049	0.049	0.049	0.049	0.051	0.051	0.053	0.053	0.069	0.075	0.076	0.097
Decision Trees	10	R2	0.785	0.787	0.788	0.788	0.791	0.787	0.795	0.782	0.795	0.768	0.777	0.628	0.609	0.348
Decision Trees	10	RMSE	0.057	0.057	0.057	0.057	0.056	0.057	0.056	0.057	0.055	0.058	0.057	0.074	0.076	0.099
Decision Trees	5	R2	0.785	0.796	0.795	0.796	0.783	0.793	0.803	0.797	0.827	0.784	0.681	0.605	0.546	0.351
Decision Trees	5	RMSE	0.058	0.057	0.057	0.057	0.058	0.057	0.055	0.056	0.052	0.058	0.070	0.078	0.083	0.100
Decision Trees	3	R2	0.772	0.771	0.773	0.772	0.772	0.770	0.786	0.778	0.736	0.781	0.629	0.638	0.527	0.366
Decision Trees	3	RMSE	0.060	0.060	0.060	0.060	0.060	0.060	0.058	0.059	0.064	0.058	0.076	0.075	0.086	0.100
Knn	20	R2	0.779	0.676	0.663	0.637	0.614	0.498	0.490	0.333	0.328	0.322	0.456	0.015	-0.208	-0.251
Knn	20	RMSE	0.055	0.068	0.069	0.072	0.075	0.086	0.087	0.099	0.100	0.100	0.089	0.119	0.132	0.134
Knn	10	R2	0.788	0.662	0.634	0.621	0.601	0.494	0.488	0.335	0.311	0.304	0.460	0.060	-0.151	-0.169
Knn	10	RMSE	0.055	0.071	0.074	0.075	0.077	0.087	0.088	0.100	0.102	0.103	0.089	0.118	0.130	0.131
Knn	5	R2	0.781	0.631	0.629	0.619	0.601	0.426	0.423	0.327	0.333	0.305	0.469	0.079	-0.095	-0.049

Table D.2: Models' performance throughout backward stepwise elimination.

Model	# Folds	Metric	It1.	It2.	It3.	It4.	It5.	It6.	It7.	It8.	It9.	It10.	It11.	It12.	It13.	It14.
Knn	5	RMSE	0.057	0.075	0.075	0.076	0.078	0.094	0.094	0.102	0.102	0.104	0.090	0.119	0.129	0.126
Knn	3	R2	0.773	0.632	0.627	0.618	0.594	0.489	0.475	0.358	0.336	0.293	0.443	0.110	-0.019	-0.002
Knn	3	RMSE	0.059	0.076	0.076	0.077	0.080	0.090	0.091	0.100	0.102	0.105	0.093	0.118	0.126	0.125
OLS	20	R2	0.202	0.202	0.165	0.164	0.194	0.186	0.179	0.179	0.170	0.169	0.129	0.068	0.051	0.025
OLS	20	RMSE	0.108	0.108	0.111	0.111	0.109	0.110	0.110	0.110	0.111	0.111	0.114	0.118	0.119	0.121
OLS	10	R2	0.198	0.198	0.166	0.165	0.203	0.194	0.187	0.187	0.178	0.177	0.133	0.069	0.053	0.027
OLS	10	RMSE	0.109	0.109	0.112	0.112	0.110	0.111	0.111	0.111	0.112	0.112	0.115	0.119	0.120	0.122
OLS	5	R2	0.222	0.224	0.187	0.187	0.210	0.202	0.195	0.195	0.187	0.184	0.133	0.076	0.059	0.032
OLS	5	RMSE	0.110	0.110	0.112	0.112	0.111	0.111	0.112	0.112	0.112	0.112	0.116	0.120	0.121	0.123
OLS	3	R2	0.217	0.215	0.177	0.177	0.212	0.201	0.198	0.197	0.191	0.187	0.128	0.061	0.045	0.035
OLS	3	RMSE	0.111	0.111	0.114	0.114	0.111	0.112	0.112	0.112	0.113	0.113	0.117	0.121	0.122	0.123
PLS	20	R2	0.187	0.209	0.210	0.212	0.211	0.221	0.218	0.208	0.209	0.196	0.175	0.154	0.110	0.045
PLS	20	RMSE	0.109	0.108	0.108	0.108	0.108	0.107	0.107	0.108	0.108	0.109	0.110	0.112	0.115	0.119
PLS	10	R2	0.189	0.211	0.212	0.213	0.213	0.224	0.220	0.212	0.212	0.200	0.180	0.161	0.122	0.054
PLS	10	RMSE	0.110	0.109	0.109	0.109	0.109	0.108	0.108	0.109	0.109	0.110	0.111	0.112	0.115	0.120
PLS	5	R2	0.211	0.224	0.226	0.228	0.227	0.238	0.234	0.227	0.225	0.215	0.193	0.177	0.130	0.066
PLS	5	RMSE	0.111	0.110	0.109	0.109	0.109	0.109	0.109	0.109	0.109	0.110	0.112	0.113	0.116	0.120
PLS	3	R2	0.214	0.232	0.233	0.235	0.234	0.237	0.236	0.230	0.230	0.228	0.206	0.189	0.142	0.074
PLS	3	RMSE	0.111	0.110	0.110	0.110	0.110	0.109	0.110	0.110	0.110	0.110	0.112	0.113	0.116	0.121
Random Forest	20	R2	0.876	0.874	0.877	0.877	0.881	0.883	0.887	0.884	0.883	0.896	0.800	0.761	0.701	0.437
Random Forest	20	RMSE	0.042	0.043	0.042	0.042	0.042	0.041	0.040	0.041	0.041	0.039	0.054	0.058	0.065	0.090
Random Forest	10	R2	0.873	0.871	0.874	0.876	0.878	0.883	0.882	0.881	0.880	0.893	0.793	0.762	0.695	0.449
Random Forest	10	RMSE	0.044	0.044	0.044	0.044	0.043	0.042	0.043	0.043	0.043	0.040	0.055	0.058	0.067	0.091
Random Forest	5	R2	0.859	0.873	0.872	0.875	0.865	0.878	0.876	0.878	0.880	0.896	0.801	0.752	0.700	0.462
Random Forest	5	RMSE	0.047	0.045	0.045	0.044	0.046	0.043	0.044	0.044	0.043	0.040	0.055	0.061	0.068	0.091
Random Forest	3	R2	0.844	0.860	0.850	0.860	0.858	0.862	0.862	0.857	0.865	0.884	0.773	0.742	0.660	0.416
Random Forest	3	RMSE	0.050	0.047	0.049	0.047	0.047	0.046	0.046	0.047	0.046	0.043	0.060	0.063	0.073	0.096
XGBoost	20	R2	0.900	0.897	0.899	0.898	0.904	0.890	0.884	0.878	0.872	0.893	0.844	0.686	0.649	0.515
XGBoost	20	RMSE	0.036	0.037	0.037	0.037	0.036	0.038	0.038	0.038	0.039	0.038	0.046	0.067	0.070	0.084
XGBoost	10	R2	0.883	0.905	0.910	0.909	0.908	0.902	0.899	0.897	0.885	0.872	0.826	0.770	0.695	0.520
XGBoost	10	RMSE	0.041	0.037	0.036	0.036	0.036	0.037	0.038	0.038	0.039	0.042	0.050	0.058	0.066	0.084
XGBoost	5	R2	0.844	0.850	0.855	0.852	0.852	0.849	0.855	0.905	0.899	0.890	0.858	0.665	0.627	0.527
XGBoost	5	RMSE	0.049	0.048	0.047	0.047	0.047	0.047	0.047	0.038	0.039	0.041	0.046	0.071	0.075	0.085
XGBoost	3	R2	0.817	0.819	0.862	0.863	0.821	0.818	0.819	0.852	0.853	0.824	0.695	0.667	0.636	0.532
XGBoost	3	RMSE	0.054	0.053	0.047	0.046	0.053	0.053	0.053	0.048	0.048	0.052	0.069	0.072	0.076	0.086

Appendix E

Excluded Features throughout the Backward Stepwise Feature Selection Method

Table E.1: Excluded features throughout stepwise backward elimination

Model	# Folds	It1.	It2.	It3.	It4.	It5.	It6.	It7.	It8.	It9.	It10.	It11.	It12.	It13.	It14.
AdaBoost	20	forecast error	moq	sales adjusted	batch size	ss	production cost	weight	pvd	lead time	#steps	material cost	machine time	bom	volume
AdaBoost	10	moq	sales adjusted	forecast error	ss	batch size	weight	lead time	pvd	volume	#steps	machine time	material cost	production cost	bom
AdaBoost	5	sales adjusted	moq	ss	forecast error	batch size	lead time	weight	pvd	volume	machine time	#steps	material cost	production cost	bom
AdaBoost	3	moq	sales adjusted	forecast error	ss	batch size	lead time	pvd	weight	#steps	machine time	production cost	bom	material cost	volume
Bayesian Reg.	20	forecast error	#steps	ss	pvd	bom	weight	batch size	machine time	production cost	lead time	material cost	moq	volume	sales adjusted
Bayesian Reg.	10	forecast error	#steps	ss	pvd	bom	weight	batch size	machine time	production cost	lead time	material cost	moq	volume	sales adjusted
Bayesian Reg.	5	forecast error	#steps	ss	pvd	bom	weight	batch size	machine time	production cost	lead time	material cost	moq	volume	sales adjusted
Bayesian Reg.	3	forecast error	#steps	ss	pvd	bom	weight	batch size	machine time	production cost	lead time	material cost	moq	volume	sales adjusted
Beta Reg.	20	forecast error	#steps	ss	pvd	bom	weight	batch size	machine time	production cost	lead time	material cost	moq	volume	sales adjusted
Beta Reg.	10	forecast error	#steps	ss	pvd	bom	weight	batch size	machine time	production cost	lead time	material cost	moq	volume	sales adjusted
Beta Reg.	5	forecast error	#steps	ss	pvd	bom	weight	batch size	machine time	production cost	lead time	material cost	moq	volume	sales adjusted
Beta Reg.	3	forecast error	#steps	ss	pvd	bom	weight	batch size	machine time	production cost	lead time	material cost	moq	volume	sales adjusted
Decision Trees	20	moq	batch size	forecast error	ss	sales adjusted	pvd	weight	production cost	lead time	#steps	material cost	bom	machine time	volume
Decision Trees	10	sales adjusted	forecast error	batch size	moq	ss	weight	pvd	production cost	lead time	bom	#steps	material cost	machine time	volume
Decision Trees	5	sales adjusted	forecast error	batch size	moq	ss	weight	pvd	production cost	lead time	#steps	material cost	bom	machine time	volume
Decision Trees	3	moq	batch size	sales adjusted	forecast error	ss	pvd	weight	production cost	lead time	#steps	material cost	bom	machine time	volume
Knn	20	#steps	material cost	production cost	machine time	weight	volume	bom	ss	sales adjusted	forecast error	pvd	batch size	moq	lead time
Knn	10	#steps	material cost	production cost	machine time	weight	volume	bom	ss	sales adjusted	forecast error	pvd	batch size	moq	lead time
Knn	5	#steps	material cost	production cost	machine time	weight	volume	bom	ss	sales adjusted	forecast error	pvd	batch size	moq	lead time
Knn	3	#steps	material cost	production cost	machine time	weight	volume	bom	ss	sales adjusted	forecast error	pvd	batch size	moq	lead time
OLS	20	forecast error	#steps	ss	pvd	bom	weight	batch size	machine time	production cost	lead time	material cost	moq	volume	sales adjusted
OLS	10	forecast error	#steps	ss	pvd	bom	weight	batch size	machine time	production cost	lead time	material cost	moq	volume	sales adjusted
OLS	5	forecast error	#steps	ss	pvd	bom	weight	batch size	machine time	production cost	lead time	material cost	moq	volume	sales adjusted
OLS	3	forecast error	#steps	ss	pvd	bom	weight	batch size	machine time	production cost	lead time	material cost	moq	volume	sales adjusted
PLS	20	pvd	batch size	ss	bom	forecast error	weight	machine time	production cost	volume	moq	sales adjusted	#steps	material cost	lead time
PLS	10	pvd	batch size	ss	bom	forecast error	weight	machine time	production cost	volume	moq	sales adjusted	#steps	material cost	lead time
PLS	5	pvd	batch size	ss	bom	forecast error	weight	machine time	production cost	volume	moq	sales adjusted	#steps	material cost	lead time
PLS	3	pvd	batch size	ss	bom	forecast error	weight	machine time	production cost	volume	moq	sales adjusted	#steps	material cost	lead time
Random Forest	20	moq	batch size	forecast error	ss	sales adjusted	pvd	production cost	weight	lead time	#steps	bom	machine time	material cost	volume
Random Forest	10	batch size	moq	forecast error	ss	sales adjusted	pvd	production cost	weight	lead time	#steps	bom	machine time	material cost	volume
Random Forest	5	moq	forecast error	batch size	ss	sales adjusted	pvd	production cost	weight	lead time	#steps	bom	machine time	material cost	volume
Random Forest	3	moq	forecast error	batch size	ss	sales adjusted	pvd	production cost	weight	lead time	#steps	bom	machine time	material cost	volume
XGBoost	20	forecast error	ss	batch size	sales adjusted	production cost	weight	volume	moq	machine time	material cost	#steps	lead time	bom	pvd
XGBoost	10	sales adjusted	forecast error	ss	batch size	production cost	weight	volume	moq	machine time	bom	material cost	lead time	#steps	pvd
XGBoost	5	forecast error	batch size	sales adjusted	ss	volume	weight	moq	production cost	machine time	material cost	#steps	lead time	bom	pvd
XGBoost	3	forecast error	batch size	ss	production cost	sales adjusted	moq	weight	volume	machine time	#steps	material cost	lead time	bom	pvd

Appendix F

Model's Performance in each Cluster

Table F.1: Models' performance metrics

Cluster	SheetName	# Folds	R2	RMSE
0	AdaBoost	3	0.734	0.112
0	AdaBoost	5	0.592	0.120
0	AdaBoost	10	0.621	0.111
0	Bayesian Reg.	3	0.269	0.169
0	Bayesian Reg.	5	0.328	0.160
0	Bayesian Reg.	10	0.370	0.151
0	Decision Trees	3	0.232	0.173
0	Decision Trees	5	0.306	0.163
0	Decision Trees	10	0.423	0.131
0	Knn	3	0.218	0.174
0	Knn	5	0.237	0.168
0	Knn	10	0.275	0.157
0	Neural Networks	3	0.041	0.194
0	PLS	3	0.009	0.197
0	Random Forest	3	0.505	0.136
0	Random Forest	5	0.533	0.131
0	Random Forest	10	0.475	0.128
0	XGBoost	3	0.475	0.139
0	XGBoost	5	0.358	0.147
0	XGBoost	10	0.463	0.122
1	AdaBoost	3	0.956	0.031
1	AdaBoost	5	0.960	0.034
1	AdaBoost	10	0.955	0.035
1	AdaBoost	20	0.959	0.032
1	Bayesian Reg.	3	0.461	0.130
1	Bayesian Reg.	5	0.495	0.126
1	Bayesian Reg.	10	0.491	0.125
1	Bayesian Reg.	20	0.441	0.123
1	Decision Trees	3	0.861	0.062
1	Decision Trees	5	0.897	0.051
1	Decision Trees	10	0.884	0.046
1	Decision Trees	20	0.906	0.032
1	Knn	3	0.789	0.080
1	Knn	5	0.830	0.071
1	Knn	10	0.834	0.069
1	Knn	20	0.821	0.063
1	Neural Networks	5	0.143	0.158
1	Neural Networks	20	0.420	0.127
1	OLS	3	0.387	0.137
1	OLS	5	0.490	0.126
1	OLS	10	0.479	0.125
1	OLS	20	0.428	0.121
1	PLS	3	0.274	0.150
1	PLS	5	0.497	0.126
1	PLS	10	0.487	0.125
1	PLS	20	0.437	0.124
1	Random Forest	3	0.886	0.056
1	Random Forest	5	0.912	0.050
1	Random Forest	10	0.904	0.048
1	Random Forest	20	0.911	0.039
1	XGBoost	3	0.903	0.051
1	XGBoost	5	0.945	0.037
1	XGBoost	10	0.926	0.038
1	XGBoost	20	0.928	0.031
2	AdaBoost	3	0.806	0.097
2	AdaBoost	5	0.728	0.077
2	AdaBoost	10	0.701	0.071
2	AdaBoost	20	0.235	0.046
2	Bayesian Reg.	3	0.200	0.200
2	Decision Trees	3	0.460	0.165
2	Decision Trees	5	0.307	0.124
2	Decision Trees	10	0.030	0.150
2	Knn	3	0.591	0.146
2	Knn	5	0.011	0.143
2	Knn	10	0.170	0.139
2	Random Forest	3	0.673	0.129
2	Random Forest	5	0.469	0.100
2	Random Forest	10	0.392	0.102
2	XGBoost	3	0.422	0.165
2	XGBoost	10	0.356	0.101

Table F.2: Model's Performance Metrics

Cluster	SheetName	# Folds	R2	RMSE
3	AdaBoost	3	0.968	0.023
3	AdaBoost	5	0.956	0.027
3	AdaBoost	10	0.964	0.021
3	AdaBoost	20	0.798	0.022
3	Bayesian Reg.	3	0.541	0.092
3	Bayesian Reg.	5	0.541	0.089
3	Decision Trees	3	0.891	0.042
3	Decision Trees	5	0.814	0.050
3	Decision Trees	10	0.809	0.041
3	Decision Trees	20	0.705	0.031
3	Knn	3	0.779	0.063
3	Knn	5	0.896	0.042
3	Knn	10	0.159	0.050
3	Neural Networks	5	0.373	0.101
3	OLS	5	0.239	0.098
3	PLS	3	0.207	0.125
3	PLS	5	0.336	0.104
3	Random Forest	3	0.913	0.039
3	Random Forest	5	0.939	0.030
3	Random Forest	10	0.943	0.028
3	Random Forest	20	0.656	0.024
3	XGBoost	3	0.920	0.039
3	XGBoost	5	0.918	0.036
3	XGBoost	10	0.840	0.039
3	XGBoost	20	0.231	0.027
4	AdaBoost	3	0.645	0.016
4	AdaBoost	5	0.674	0.015
4	AdaBoost	20	0.745	0.011
4	Bayesian Reg.	3	0.541	0.018
4	Bayesian Reg.	5	0.494	0.018
4	Bayesian Reg.	10	0.530	0.016
4	Bayesian Reg.	20	0.421	0.015
4	Decision Trees	3	0.591	0.017
4	Decision Trees	10	0.671	0.014
4	Knn	3	0.642	0.016
4	Knn	5	0.738	0.014
4	Knn	10	0.787	0.012
4	Knn	20	0.826	0.009
4	OLS	3	0.674	0.015
4	OLS	5	0.648	0.016
4	OLS	10	0.696	0.014
4	OLS	20	0.671	0.012
4	PLS	3	0.415	0.020
4	PLS	5	0.429	0.020
4	PLS	10	0.449	0.018
4	PLS	20	0.455	0.017
4	Random Forest	3	0.656	0.015
4	Random Forest	5	0.635	0.016
4	Random Forest	10	0.661	0.013
4	Random Forest	20	0.538	0.010
4	XGBoost	3	0.710	0.014
4	XGBoost	5	0.765	0.013
4	XGBoost	10	0.849	0.009
4	XGBoost	20	0.885	0.008
5	AdaBoost	3	0.920	0.017
5	AdaBoost	5	0.912	0.018
5	AdaBoost	10	0.907	0.016
5	AdaBoost	20	0.930	0.012
5	Bayesian Reg.	3	0.757	0.036
5	Bayesian Reg.	5	0.769	0.035
5	Bayesian Reg.	10	0.720	0.035
5	Bayesian Reg.	20	0.689	0.035
5	Decision Trees	3	0.831	0.030
5	Decision Trees	5	0.870	0.024
5	Decision Trees	10	0.868	0.019
5	Decision Trees	20	0.892	0.014
5	Knn	3	0.872	0.025
5	Knn	5	0.894	0.023
5	Knn	10	0.856	0.024
5	Knn	20	0.840	0.023

Table F.3: Model's Performance Metrics

Cluster	SheetName	# Folds	R2	RMSE
5	Neural Networks	20	0.140	0.053
5	OLS	3	0.626	0.042
5	OLS	5	0.480	0.043
5	PLS	3	0.708	0.040
5	PLS	5	0.702	0.040
5	PLS	10	0.657	0.040
5	PLS	20	0.628	0.039
5	Random Forest	3	0.907	0.021
5	Random Forest	5	0.905	0.020
5	Random Forest	10	0.915	0.016
5	Random Forest	20	0.913	0.015
5	XGBoost	3	0.921	0.019
5	XGBoost	5	0.933	0.015
5	XGBoost	10	0.934	0.013
5	XGBoost	20	0.939	0.012
6	AdaBoost	3	0.832	0.022
6	AdaBoost	5	0.829	0.022
6	AdaBoost	10	0.828	0.022
6	AdaBoost	20	0.825	0.022
6	Bayesian Reg.	3	0.197	0.049
6	Bayesian Reg.	5	0.184	0.049
6	Bayesian Reg.	10	0.181	0.049
6	Bayesian Reg.	20	0.161	0.049
6	Decision Trees	3	0.622	0.033
6	Decision Trees	5	0.655	0.032
6	Decision Trees	10	0.715	0.027
6	Decision Trees	20	0.730	0.024
6	Knn	3	0.644	0.033
6	Knn	5	0.706	0.029
6	Knn	10	0.738	0.027
6	Knn	20	0.739	0.026
6	PLS	3	0.020	0.054
6	PLS	5	0.005	0.055
6	Random Forest	3	0.808	0.024
6	Random Forest	5	0.828	0.023
6	Random Forest	10	0.857	0.020
6	Random Forest	20	0.843	0.020
6	XGBoost	3	0.838	0.022
6	XGBoost	5	0.859	0.020
6	XGBoost	10	0.899	0.017
6	XGBoost	20	0.894	0.016
7	Knn	3	0.412	0.064
7	Knn	5	0.573	0.054
8	AdaBoost	3	0.758	0.018
8	AdaBoost	5	0.721	0.019
8	AdaBoost	10	0.776	0.017
8	AdaBoost	20	0.747	0.017
8	Bayesian Reg.	3	0.310	0.030
8	Bayesian Reg.	5	0.350	0.029
8	Bayesian Reg.	10	0.348	0.029
8	Bayesian Reg.	20	0.361	0.028
8	Decision Trees	3	0.789	0.016
8	Decision Trees	5	0.748	0.017
8	Decision Trees	10	0.726	0.018
8	Decision Trees	20	0.790	0.014
8	Knn	3	0.723	0.019
8	Knn	5	0.714	0.019
8	Knn	10	0.768	0.017
8	Knn	20	0.810	0.014
8	OLS	3	0.481	0.026
8	OLS	5	0.502	0.025
8	OLS	10	0.470	0.026
8	OLS	20	0.505	0.025
8	Random Forest	3	0.785	0.017
8	Random Forest	5	0.822	0.015
8	Random Forest	10	0.846	0.013
8	Random Forest	20	0.887	0.010
8	XGBoost	3	0.763	0.017
8	XGBoost	5	0.790	0.016
8	XGBoost	10	0.811	0.014
8	XGBoost	20	0.865	0.011

Appendix G

Theoretical Flow Efficiency vs Predicted Flow Efficiency

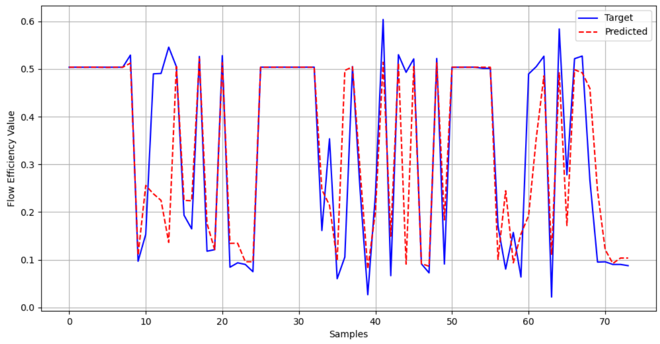


Figure G.1: Theoretical flow efficiency values vs Predicted flow efficiency values for cluster 0 with XGBoost

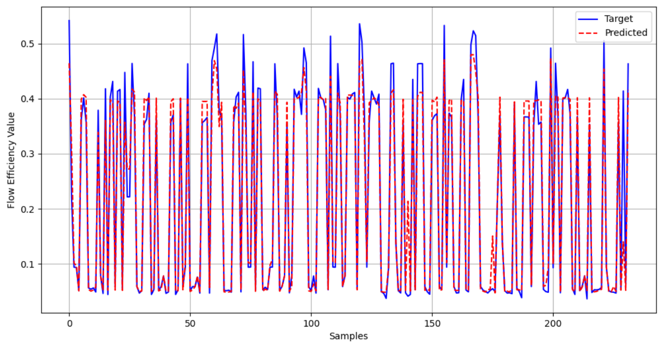


Figure G.2: Theoretical flow efficiency values vs Predicted flow efficiency values for cluster 1 with K-nn

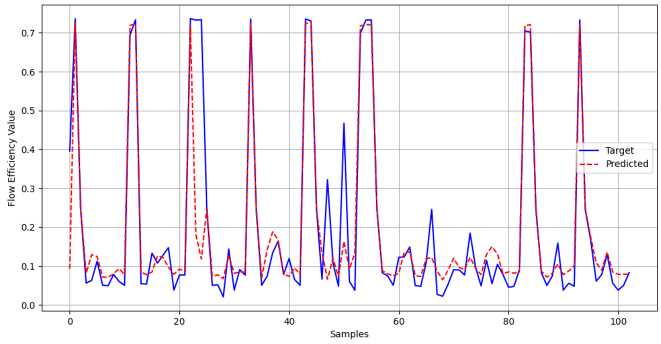


Figure G.3: Theoretical flow efficiency values vs Predicted flow efficiency values for cluster 2 with OLS

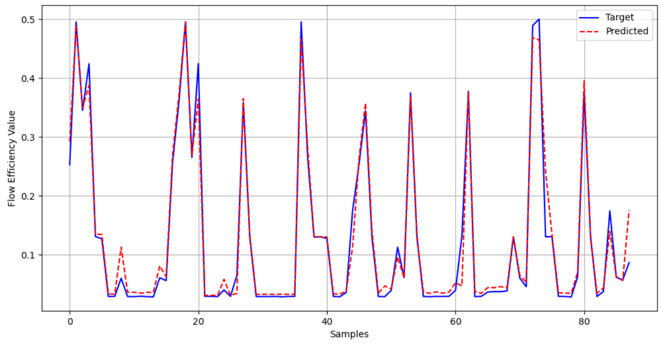


Figure G.4: Theoretical flow efficiency values vs Predicted flow efficiency values for cluster 3 with Random Forest

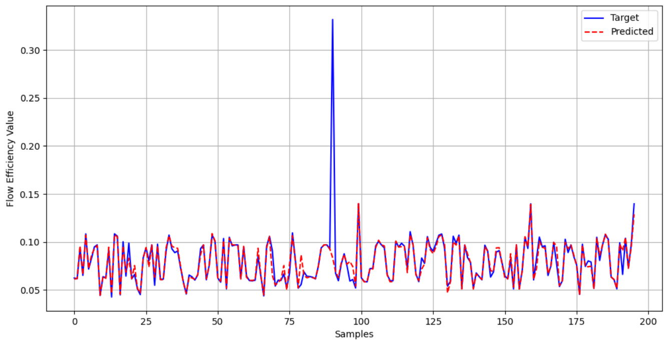


Figure G.5: Theoretical flow efficiency values vs Predicted flow efficiency values for cluster 4 with XGBoost

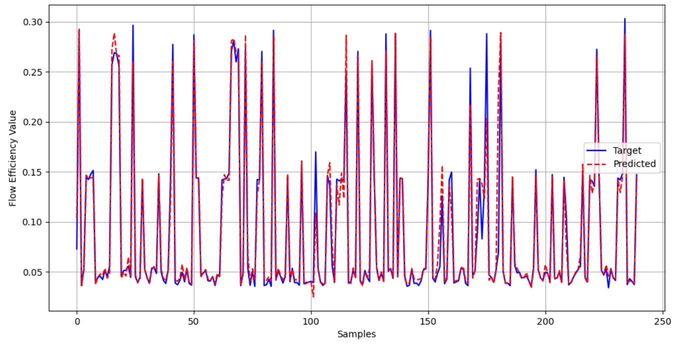


Figure G.6: Theoretical flow efficiency values vs Predicted flow efficiency values for cluster 5 with Random Forest

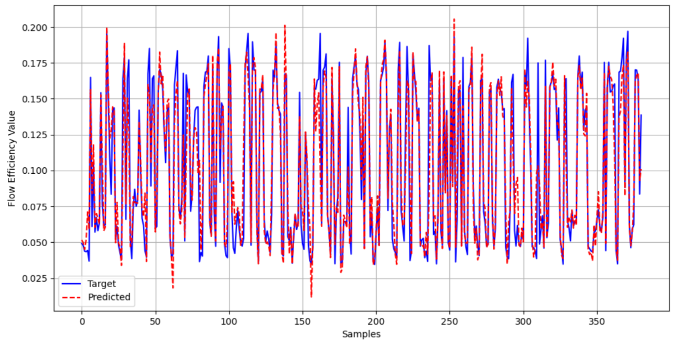


Figure G.7: Theoretical flow efficiency values vs Predicted flow efficiency values for cluster 6 with Decision Trees

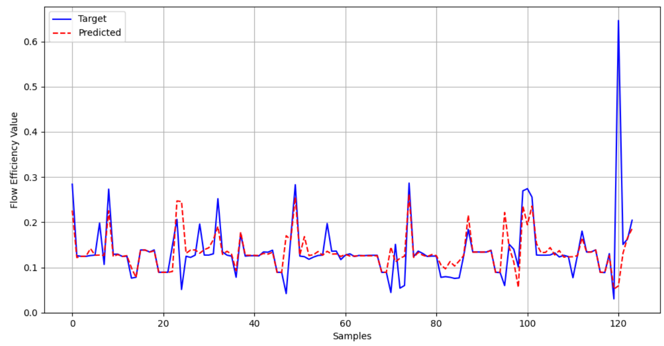


Figure G.8: Theoretical flow efficiency values vs Predicted flow efficiency values for cluster 7 with K-nn

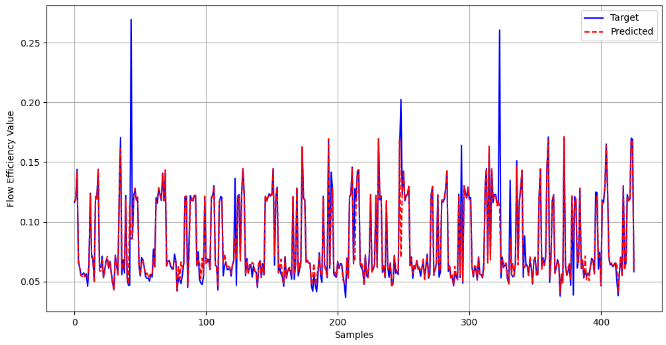


Figure G.9: Theoretical flow efficiency values vs Predicted flow efficiency values for cluster 8 with AdaBoost

Appendix H

Normality Distribution Test Results

Table H.1: Normality distribution test of data for each flow efficiency driver in each one of the identified clusters

	#steps		material cost		production cost		machine time		weight		volume		bom	
Cluster	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value
0	0.368982	3.19E-16	0.933977	0.000796	0.916979	0.000128	0.795728	9.48E-09	0.840799	1.91E-07	0.651730	5.80E-12	0.715481	1.12E-10
1	1	1	0.873341	5.67E-13	0.543824	3.27E-24	0.922065	1.07E-09	0.880062	1.41E-12	0.961917	7.54E-06	0.891580	7.26E-12
2	0.659733	4.12E-14	0.931177	4.12E-05	0.788278	7.18E-11	0.719204	9.65E-13	0.733655	2.23E-12	0.784079	5.38E-11	0.760040	1.11E-11
3	1	1	0.651142	3.79E-13	0.773928	2.60E-10	0.586786	2.32E-14	0.773302	2.50E-10	0.678627	1.40E-12	0.699629	3.99E-12
4	1	1	0.906792	9.28E-10	0.822741	3.52E-14	0.376967	1.59E-25	0.899576	3.18E-10	0.714738	4.71E-18	0.689314	8.38E-19
5	1	1	0.669773	2.27E-21	0.608736	4.95E-23	0.441975	9.98E-27	0.810985	2.26E-16	0.428078	5.40E-27	0.760551	2.11E-18
6	1	1	0.960068	1.15E-08	0.926007	8.80E-13	0.981028	6.64E-05	0.949373	3.84E-10	0.966528	1.18E-07	0.987021	0.001751
7	0.884865	2.38E-08	0.883382	2.02E-08	0.886390	2.82E-08	0.895480	7.90E-08	0.868513	4.25E-09	0.855398	1.18E-09	0.898845	1.17E-07
8	0.475170	1.72E-33	0.900322	4.55E-16	0.815592	9.85E-22	0.843718	4.06E-20	0.772475	7.03E-24	0.674519	8.70E-28	0.502153	7.83E-33

Table H.2: Normality distribution test of data for each flow efficiency driver in each one of the identified clusters

	safety stock		sales		forecast error		pvd		batch size		moq		lead time	
Cluster	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	Statistic	P-Value
0	0.710317	8.66E-11	0.917049	0.000128	0.958702	0.016427	0.965173	0.039030	0.583972	3.67E-13	0.295073	4.28E-17	0.893225	1.30E-05
1	0.982377	0.005521	0.920709	8.33E-10	0.910332	1.37E-10	0.878164	1.08E-12	0.711442	9.25E-20	1	1	0.922388	1.13E-09
2	0.989227	0.581373	0.934297	6.89E-05	0.879406	1.22E-07	0.799990	1.63E-10	0.884161	1.97E-07	0.073529	1.63E-22	0.713664	7.06E-13
3	0.464047	2.44E-16	0.587055	2.35E-14	0.864522	1.84E-07	0.815571	4.10E-09	0.600563	4.11E-14	0.954220	0.003519	0.709991	6.83E-12
4	0.890619	8.99E-11	0.780156	7.19E-16	0.951041	2.95E-06	0.977500	0.003044	0.869293	5.72E-12	0.045866	3.46E-30	0.957901	1.44E-05
5	0.655242	8.72E-22	0.474320	4.38E-26	0.340866	1.46E-28	0.405040	2.00E-27	0.686534	7.14E-21	0.038140	5.81E-33	0.346405	1.82E-28
6	0.956842	3.91E-09	0.906885	1.50E-14	0.904617	9.66E-15	0.911381	3.70E-14	0.900361	4.29E-15	0.249806	1.95E-36	0.851169	1.50E-18
7	3.00E-12	0.667918	2.30E-15	0.852506	8.94E-10	0.778640	2.15E-12	0.967839	0.004692	0.293051	7.62E-22	0.767304	9.66E-13	
8	0.147658	7.48E-40	0.274159	1.11E-37	0.717668	3.38E-26	0.832521	8.73E-21	0.479485	2.18E-33	0.273595	1.08E-37	0.617177	1.16E-29