ANALYZE FACTORS AFFECTING THREAD CONSUMPTION IN A GARMENT AND DEVELOP A MACHINE LEARNING-BASED PREDICTION MODEL

V.W. Ubayawickrema

2024

Analyze Factors Affecting Thread Consumption in a Garment and Develop a Machine Learning-based Prediction Model

A dissertation submitted for the Degree of Master of Business Analytics

V.W. Ubayawickrema
University of Colombo School of Computing
2024

DECLARATION

Name of the student: Vindya Ubayawickrema
Registration number: 2020/BA/038
Name of the Degree Programme: Master of Business Analytics
Project/Thesis title: Analyze Factors Affecting Thread Consumption in a Garment and Develop a Machine Learning-based Prediction Model

1. The project/thesis is my original work and has not been submitted previously for a degree at this or any other University/Institute. To the best of my knowledge, it does not contain any material published or written by another person, except as acknowledged in the text.

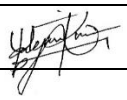
2. I understand what plagiarism is, the various types of plagiarism, how to avoid it, what my resources are, who can help me if I am unsure about a research or plagiarism issue, as well as what the consequences are at University of Colombo School of Computing (UCSC) for plagiarism.

3. I understand that ignorance is not an excuse for plagiarism and that I am responsible for clarifying, asking questions and utilizing all available resources in order to educate myself and prevent myself from plagiarizing.

4. I am also aware of the dangers of using online plagiarism checkers and sites that offer essays for sale. I understand that if I use these resources, I am solely responsible for the consequences of my actions.

5. I assure that any work I submit with my name on it will reflect my own ideas and effort. I will properly cite all material that is not my own.

6. I understand that there is no acceptable excuse for committing plagiarism and that doing so is a violation of the Student Code of Conduct.

Signature of the Student	**Date (DD/MM/YYYY)**
	25/09/2024

Certified by Supervisor(s)

This is to certify that this project/thesis is based on the work of the above-mentioned student under my/our supervision. The thesis has been prepared according to the format stipulated and is of an acceptable standard.

	Supervisor 1	**Supervisor 2**	**Supervisor 3**
Name	Dr. Samantha Mathara Arachchi		
Signature			
Date	25/09/2024		

ACKNOWLEDGEMENTS

First and foremost, I would like to express my utmost gratitude and appreciation to my internal supervisor DR. S. S. P. Mathara Arachchi for his invaluable guidance and assistance given to me throughout this research study. Furthermore, I am extending my heartfelt thanks to my industry supervisor Ms. Harshana Rodrigo of Emjay International (Pvt) Ltd for her continuous support and encouragement to complete this study successfully.

I wish to express my sincere appreciation to Mr. R.J. Amaraweera, the course coordinator, all the senior lecturers, lecturers, and instructors of the University of Colombo School of Computing for the guidance and support they have provided.

Further, I would like to acknowledge the management of Emjay International (Pvt) Ltd. For the fullest corporation in helping me conduct the project and guiding me with advice and direction to compile a valuable report.

Words fail to express my gratitude towards my family for their love, support and guidance. There have been many sacrifices from my family to help me achieve my academic goals. Especially in completing my Masters, with the many challenges that we faced together.

Finally, I would like to extend my sincere thanks to my friends and everyone who supported me, guided me, and encouraged me throughout my time at the university. I’m forever grateful!

ABSTRACT

The garment manufacturing industry faces intensified competition, prompting the need for cost control and efficient inventory management. This research addresses the challenges of excess thread stock, leading to increased write-off expenses and environmental concerns. Focusing on predicting sewing thread consumption in underwear fullbrief styles, the study employs statistical and machine learning techniques, considering variables such as garment style, fabric/seam thickness, stitch length, stitch density/SPI, seam type, and estimated wastage.

The development of a user interface using Streamlit integrates machine learning models for two types of threads, allowing users to input parameters through an intuitive layout. The user-friendly interface facilitates informed decision-making based on predictions of total thread consumption. The application contributes to reducing write-off expenses, minimizing inventory costs, and aligning with environmental sustainability goals.

The research highlights the effectiveness of machine learning models, particularly artificial neural network models, in predicting thread consumption. Overcoming challenges such as overfitting and enhancing generalization, the study emphasizes the need for refining model architectures and exploring additional features. The user interface development emerges as a crucial tool for achieving efficient cost control and sustainability in the garment manufacturing industry.

Keywords: Garment Manufacturing, Thread Consumption Prediction, Machine Learning, Artificial Neural Network, User Interface, Streamlit, Cost Control.

TABLE OF CONTENTS

DECLARATION i

ACKNOWLEDGEMENTS ii

ABSTRACT iii

LIST OF FIGURES vii

LIST OF TABLES viii

LIST OF APPENDICES ix

LIST OF ABBREVIATIONS x

CHAPTER 1 INTRODUCTION 1

1.1 Motivation 3

1.2 Research Problem 3

1.3 Research Aims And Objectives 4

1.3.1 Aims Of The Study 4

1.3.2 Objectives Of The Study 4

1.4 Research Questions 4

1.5 Scope Of Work 5

1.6 Organization Of The Study 7

CHAPTER 2 LITERATURE REVIEW 8

2.1 Factors Affecting Thread Consumption 8

2.2 Graphs, Tables And Formulas 10

2.3 Mathematical And Geometrical Models 12

2.4 Machine Learning, Neural Network Models And Metaheuristic Optimization Models 14

2.5 Chapter Summary 17

CHAPTER 3 METHODOLOGY AND THEORY 18

3.1 Methodology 18

3.1.1 Data Collection 18

3.1.2 Data Preparation And Preprocessing 20

3.1.3 Model Training 25

3.1.4 Model Evaluation and Selection 26
3.1.5 Deployment 26
3.2 Theory 27
3.2.1 Statistical Models 27
3.2.2 Machine Learning Models 28
3.2.3 Model Comparison 33
3.3 Research/Solution Design 34
3.4 Chapter Summary 35
CHAPTER 4 EXPLORATORY DATA ANALYSIS 36
4.1 Buffer Wastage Inbuilt By The Product Development Vs Measured Wastage 36
4.2 Influence Of Spi On The Consumed Thread Amount Behavior 37
4.3 Influence Of Seam Length On The Consumed Thread Amount Behavior 38
4.4 Influence Of Seam Thickness On The Consumed Thread Amount Behavior 38
4.5 Influence Of Wastage On The Consumed Thread Amount Behavior 39
4.6 Influence Of Operator Skill On The Consumed Thread Amount Behavior 40
4.7 Chapter Summary 40
CHAPTER 5 EVALUATION AND RESULTS 41
5.1 Multivariate Linear Regression 41
5.2 Ensemble Models 44
5.2.1 Random Forest 44
5.2.2 Gradient Boosting 46
5.2.3 XGBoost 48
5.3 Multilayer Perceptron (Neural Network) 50
5.4 Development Of A User Interface 53
5.5 Chapter Summary 55
CHAPTER 6 DISCUSSION, FUTURE WORK AND CONCLUSION 56
6.1 Discussion 56
6.2 Limitations 60

6.3 Suggestions For Future Work 61
6.4 Conclusion 61
LIST OF REFERENCES I
APPENDICES IV

LIST OF FIGURES

Figure 1.1 Inventory Write-Off for Financial Year 2022/23 1
Figure 1.2 Different types of Stitches used in the study 6

Figure 2.1 AMANN and A&E sewing thread requirement tables 11
Figure 2.2 Relationship between the waste factor and experimental consumption values 14
Figure 2.3 Post-training analysis of neural network results 16

Figure 3.1 Proposed Methodology 18
Figure 3.2 Data Collection Process Flow Diagram 19
Figure 3.3 PD's Thread Consumption Worksheet 20
Figure 3.4 Harmonized Data Set 21
Figure 3.5 Feature Engineered Data set 24
Figure 3.6 Artificial Intelligence 29
Figure 3.7 Framework of an Artificial Neural Network 32

Figure 4.1 Buffer Wastage inbuilt by the Product Development vs Measured Wastage 37
Figure 4.2 Influence of SPI on the thread consumption 37
Figure 4.3 Influence of Seam Length on the thread consumption 38
Figure 4.4 Correlations between Seam Thickness and thread consumption 39
Figure 4.5 Correlation between Wastage and thread consumption 39
Figure 4.6 Influence of Operator Skill on the thread consumption 40

Figure 5.1 Mutual Information Scores Plot for Linear Regression models 42
Figure 5.2 Actual vs Predicted values of Linear Regression model 43
Figure 5.3 Mutual Information Scores Plot for Random Forest model 44
Figure 5.4 Actual vs Predicted values of Random Forest model 46
Figure 5.5 Actual vs Predicted values of Gradient Boosting model 47
Figure 5.6 Feature Importance Plot for XGBoost model 48
Figure 5.7 Actual vs Predicted values of XGBoost model 49
Figure 5.8 Actual vs Predicted values of ANN model 52
Figure 5.9 Proposed Interface for predicting thread consumption 53
Figure 5.10 Expanded Operation - User Input Interface 54

LIST OF TABLES

Table 3.1 Dataset Description 21
Table 3.2 Architectural Decisions & Rationale 34

Table 5.1 Evaluation matrices of Linear Regression Models - Wastage Model 42
Table 5.2 Evaluation matrices of Linear Regression Models - Consumption Model 43
Table 5.3 Evaluation matrices of Random Forest model - Wastage Model 45
Table 5.4 Evaluation matrices of Random Forest model - Consumption Model 45
Table 5.5 Evaluation matrices of Gradient Boosting model - Wastage Model 47
Table 5.6 Evaluation matrices of Gradient Boosting model - Consumption Model 47
Table 5.7 Evaluation matrices of XGBoost model - Wastage Model 48
Table 5.8 Evaluation matrices of XGBoost model - Consumption Model 49
Table 5.9 Consolidated Evaluation matrices of Ensemble model – Consumption Model 50
Table 5.10 Evaluation matrices of the ANN Models - Wastage Model 51
Table 5.11 Evaluation matrices of the ANN Models - Consumption Model 51
Table 5.12 Evaluation matrices of the ANN Wastage model pre and post feature selection 51
Table 5.13 Evaluation matrices of the ANN Consumption model pre and post feature selection 52

Table 6.1 Comparison of different techniques of thread consumption prediction 59

LIST OF APPENDICES

Appendix A : Supported Documents.....IV
Appendix B: Samples of the Python Code V
Appendix C: Drafted Research Paper..... XII

LIST OF ABBREVIATIONS

A&E	American & Efrid
ACO	Ant Colony Optimization
ANN	Artificial Neural Network
B2B	Business-To-Business
BP	Business Process
CM	Contribution Margin
COL	Color
DL	Deep Learning
EDA	Exploratory Data Analysis
ERP	Enterprise Resource Planning
GA	Genetic Algorithm
L	Looper
M/C Aallo	Machine Allocation
MAE	mean absolute error
MI	Mutual information
ML	Machine Learning
MLP	Multilayer Perceptron
MO	Machine Operator
N	Needle
NF	Knitted Fabric
OB	Operation Breakdown
OTD	On-Time Delivery
PD	Product Development
PSO	Particular Swarm Optimization
RMSE	root mean square error
SPI	Stitches Per Inch
TKT	Ticket
U.S.A	United States of America
XGB	Extreme Gradient Boosting

CHAPTER 1

INTRODUCTION

The garment manufacturing industry is currently experiencing increased competition, which drives organizations to produce garments at the lowest possible cost (Sharma , et al., 2017). Therefore there is a need for strict cost control from the point of ordering raw materials to completing the orders. Raw material cost in the direct cost portion is the fundamental cost factor affecting the contribution margin (CM margin %). The raw material cost portion includes the cost of the material that is actually utilized in the production, the material that is wasted during the production and the material stock that remained unused at the completion of the order.

The total leftover material at the author's organization amounts to $ 1,479,457, with fabric accounting for the majority of this amount ($1,023,276) and sewing thread coming in second ($75,059) (Emjay International (Pvt) Ltd., 2023). At the author's organization, this remaining stock at the order completion is referred to as write-off stock. Write-off primarily refers to a business accounting expense reported to account for losses on inventory value, thus leading to a lower profit. Currently, upon completion of the production process, the author's organization experiences a write-off of 4.1% of the materials that were originally allotted for the orders as shown in Figure 1.1 (Emjay International (Pvt) Ltd., 2023). This amount surpasses the company standard which dictates that the write-off rate should not be higher than 2% (Emjay International (Pvt) Ltd., 2023).

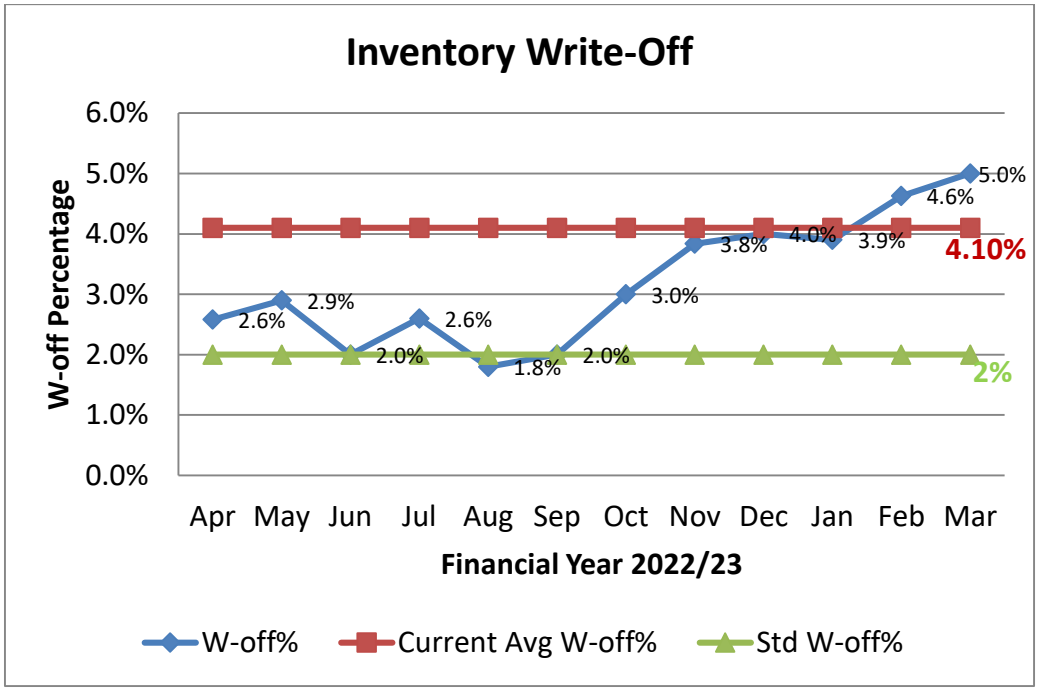


Figure 1.1 Inventory Write-Off for Financial Year 2022/23

This discrepancy points out a critical issue as it increases production costs, lowers profitability and also causes an accumulation of an excess inventory. Therefore it is imperative to address and find ways to minimize the write-off percentage and improve the overall profitability of the organization.

As an initiative, in this study, the author mainly focuses on minimizing the leftover thread stock valued at $75,059 which was not utilized during the production (Emjay International (Pvt) Ltd., 2023). The reason to consider thread leftovers against fabric leftovers is that thread leftover stock arising from human errors, mistakes or the quality of the thread is significantly minimal whereas fabric leftovers often arise from damages, errors and mistakes done by the machine operators and the poor quality of the fabric received from suppliers. Therefore it is ideal to consider only the thread leftovers initially for the prediction model to be employed at the company so that the impact from the human-dependent factors and quality of the material is kept at a minimum.

For this purpose, it is crucial to review the initial steps of determining the thread consumption thoroughly, especially given the considerable volume of stock that remains unused after the production process. Enhancing the accuracy in determining the required sewing thread consumption can lead to optimum quantities of thread needed for garment manufacturing avoiding any unused stock getting accumulated in the production floor.

The amount of thread used in the manufacturing of garments varies not just across various styles of garments but even within the garments of the same style. Variations in sizes, styles and the fabric used in the garment heavily influence the amount of thread required to complete a garment. Furthermore, stitch length, stitch density and seam type all affect thread consumption (Ukponmwan, et al., 2000). Therefore to analyze these factors and precisely predict the amount of thread that will be used in a garment, a machine learning-based model will be developed and validated as part of this study. This will enable more effective resource utilization and a decrease in the amount of leftover thread stock thus contributing to the bottom line of the organization.

1.1 Motivation

The heightened competition in the garment manufacturing sector necessitates a meticulous approach to minimize production costs and improve overall profitability. The author's organization currently faces a significant issue with a write-off rate of 4.1% on materials, surpassing the company standard of 2% (Emjay International (Pvt) Ltd., 2023). This discrepancy results in increased production costs, decreased profitability, and excess inventory accumulation. By specifically focusing on minimizing leftover thread stock, valued at $75,059, the research aims to enhance accuracy in determining thread consumption (Emjay International (Pvt) Ltd., 2023).

Existing methods for estimating thread consumption, such as graphs, tables, and formulas, often lack flexibility and accuracy, leading to variations in predictions. The literature also explores mathematical, geometrical, and machine learning-based models, showcasing the limitations of traditional techniques and the superior performance of machine learning in predicting thread consumption. Developing a machine learning-based prediction model will enable a comprehensive analysis of various factors influencing thread usage, such as garment styles, sizes, fabric types, stitch length, density, and seam types. Ultimately, the research seeks to accurately predict thread quantities required for garment manufacturing, reduce unused stock, and contribute positively to the organization's bottom line.

1.2 Research Problem

The problem identified here is the accumulation of leftover sewing thread stock at the completion of the garment manufacturing process, implying potential overestimation of the required thread consumption in a garment than what is actually required in the manufacturing process, which ultimately affects various aspects of the organization. Listed below are some of the aspects.

I. Increased inventory costs as the company needs to store and manage the leftover stock. This additional inventory management results in additional costs, including handling, storage and transport charges which have an immediate effect on the company’s profitability.
II. The company’s sustainability goals are affected since the company needs to dispose the leftover thread stock which could have an adverse impact on the environment in terms of waste generation, pollution and increased carbon footprint.

1.3 Research Aims And Objectives

The aims and objectives of conducting this study are presented in this section.

1.3.1 Aims Of The Study

The aim of this research is to analyze the various factors influencing sewing thread consumption in garment manufacturing and develop a precise machine learning-based prediction model. By exploring garment-related variables such as type, size, design, and fabric characteristics, along with stitch parameters, the goal is to improve the accuracy of estimating sewing thread requirements. Through the creation and validation of a machine learning model, this study aims to offer a practical solution for estimating thread consumption, addressing the challenges of exceeding write-off rates, and excess inventory accumulation, ultimately contributing to enhanced cost control and improved profitability.

1.3.2 Objectives Of The Study

The main objective of the study is to develop a model using machine learning techniques to predict sewing thread consumption for two thread types, across various sewing operations. The main objective is broken down as follows.

1. Analyzing the key factors affecting thread consumption in a garment using regression analysis and exploratory data analysis.
2. Applying different machine learning models to predict the thread consumption in a garment accurately.
3. Evaluating the accuracy of the applied machine learning models.
4. Developing a user friendly interface for the prediction of thread consumption.

Upon successful completion of the project, above objectives are expected to be accomplished.

1.4 Research Questions

In this section, a set of research questions tailored to each objective mentioned under Section 1.3.2 was presented, facilitating a comprehensive exploration of the factors influencing thread consumption and the effectiveness of machine learning models in prediction. The research questions are as follows.

1. What are the primary factors influencing thread consumption in a garment production process?
2. Which machine learning models are to be applied in predicting thread consumption?

3. How does the choice of features influence the performance of machine learning models in predicting thread consumption?
4. What metrics are most appropriate for evaluating the accuracy of machine learning models in predicting thread consumption and are there specific machine learning models that outperform others in predicting thread consumption?

1.5 Scope Of Work

This study primarily focuses on developing a prediction model that can forecast the actual thread consumed for a certain garment style in the organization Emjay International (Pvt) Ltd., with a higher accuracy, using machine learning techniques. A higher accuracy level is considered for the study since if the accuracy is less, the amount of thread required for the production would have a significant difference between what is predicted and actual thread consumption. This could lead to under or overutilization of resources, thus increasing production costs and eventually having an impact on profitability. Moreover, inaccurate predictions may also affect on-time delivery (OTD), leading to delays or longer lead times which may reduce the organization's ability to compete in the market and customer dissatisfaction. This will evaluate historical data from the company's ERP system and individual style-wise worksheets to discover patterns and trends in thread utilization.

This study centers its attention on the prediction of thread consumption within the specific product category of underwear fullbrief styles with the size Medium. It is noteworthy that this research has evolved from its initial proposal, which outlined the inclusion of two distinct product categories, encompassing both underwear and shirts. The decision to narrow the research scope to a single product category is due to the intricacies and nuances associated with thread consumption prediction within each product category warrant dedicated and in-depth analysis. By focusing exclusively on underwear fullbrief styles, we are afforded the opportunity to explore these complexities comprehensively, ensuring a thorough examination of all relevant factors. Out of several underwear product categories, fullbrief styles have been chosen based on their high production volume and simple construction with less number of operations to complete the garment. Construction of a fullbrief usually includes operations such as;

- Front and back gusset attach
- Elastic attach to leg
- Elastic attach to back waist
- Side seam with label
- Side tack x 4

- Cut lace
- Lace attach to front waist

The study will also concentrate on predicting the thread consumption in a garment for three commonly used stitch types for a particular fullbrief style. The 3 stitch types that will be evaluated are 301 (lockstitch), 406 (coverseam) and 514(four thread overedge). This research represents a modification from its original proposal, which outlined a broader spectrum of five stitch types, including Lockstitch, Overlock, Chainstitch, Bar tack, and Flat seam. In the course of the investigation, it became evident that fullbrief styles were inclusive of only below shown 3 stitch types (Figure 1.2) with 2 different thread types as TKT 120 and TKT 160. TKT (Ticket) numbers are a numerical system used to measure the linear density or thickness of the thread. The TKT number indicates the length of thread (in thousands of yards) that would weigh one pound. Higher TKT numbers generally represent finer or thinner threads, while lower TKT numbers indicate thicker threads. Also the thread consumed for each stitch can be varied within the same stitch type due to the differences in needle widths in each stitch.

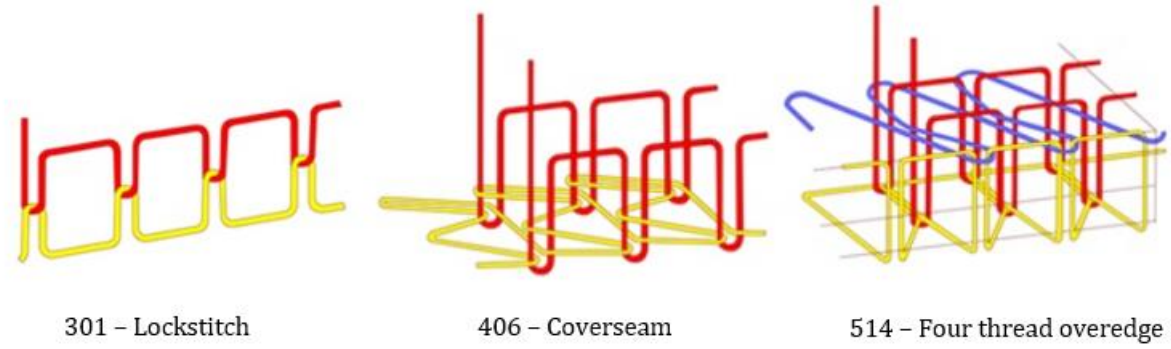


Figure 1.2 Different types of Stitches used in the study

The proposed model can be enhanced to include all the product categories manufactured at Emjay taking all the stitch types into consideration. This will allow a more heuristic understanding of actual thread consumption trends and make it easier to spot opportunities for cost and time saving throughout the whole product range. Nevertheless, doing so will necessitate collecting more data and training complex models.

1.6 Organization Of The Study

The thesis is organized as follows.

Chapter 1: Introduces the background, problem definition, and scope of the study, along with the objectives of the study.

Chapter 2: Provides a clear and concise review of literature based on past studies on predicting the amount of thread consumed in a garment.

Chapter 3: Describes the methodological background which includes a description regarding the source of the data, variables of interest, and pre-processing of the datasets and the theories applicable to the study

Chapter 4: Presents the outputs of the descriptive analysis conducted on the selected variables

Chapter 5: Provides the steps carried out when accomplishing the study's objectives and the final results.

Chapter 6: Includes the discussion on the results obtained, the issues encountered, and the limitations of the study and provides suggestions for future work. The chapter ends with stating the conclusions arrived at from the study.

CHAPTER 2

LITERATURE REVIEW

This section outlines related research initiatives that are similar to the author's study in examining the factors influencing sewing thread consumption in a garment and building up a machine learning-based model.

Competition in the apparel industry is steadily rising, leading firms to prioritize cost efficiency (DOĞAN & PAMUK, 2014). They strive to minimize total costs by enhancing the supply process, expanding purchasing options, and reducing material expenses in response to increased competition. Maintaining optimal cost levels is crucial for firms to improve their competitive advantage. Consequently, accurately calculating the materials used in production is vital for cost control and ensuring a smooth workflow. Determining the precise amount of sewing thread consumed in an apparel product holds great importance for these reasons (DOĞAN & PAMUK, 2014).

2.1 Factors Affecting Thread Consumption

The quantities of sewing thread used can vary based on the type of clothing as well as the size, design, and fabric of the same type of garment (Ukponmwan, et al., 2000). Alongside these factors, the amount of thread used is also influenced by stitch length, stitch density, and the type of stitch employed. A more comprehensive elucidation of how every aspect influences the quantity of sewing thread consumption is given below.

- Garment type: The amount of thread used will vary depending on the type of garment being sewn. For example, a trouser will use more thread than an underwear such as a boxer, and a dress will use even more thread.
- Size: The size of the garment will also affect the amount of thread used. A larger garment will use more thread than a smaller garment.
- Design: The design of the garment can also affect the amount of thread used. For example, a garment with a lot of details will use more thread than a garment with a simple design.
- Fabric: The type of fabric used in the garment can also affect the amount of thread used. Heavier fabrics will use more thread than lighter fabrics.
- Stitch length: The stitch length is the distance between each stitch. A longer stitch length will use more thread than a shorter stitch length.

- Stitch density: The stitch density is the number of stitches per inch (SPI). A higher stitch density will use more thread than a lower stitch density.
- Stitch type: The stitch type is the type of stitch used to sew the garment. Some stitch types, such as overlock stitches, use more thread than other stitch types, such as straight stitches.

It is important to consider all of these factors when estimating the amount of sewing thread needed for a garment (Ukponmwan, et al., 2000) . By understanding how these factors affect the amount of thread used, you can ensure that you have enough thread on hand to complete the project.

Yeşilpınar and Alkiraz (2005) investigated the effect of fabric thickness on sewing thread consumption. They analyzed 10 different woven fabrics with varying thicknesses and sewed each fabric with 4 different stitch types: lockstitch, chain stitch, three-thread overedge stitch, and four-thread overedge stitch. The results of the study showed that fabric thickness has a significant impact on sewing thread consumption. The amount of sewing thread consumed increased directly with the increase in fabric thickness (Yeşilpınar & Alkiraz, 2005). This is because thicker fabrics require more thread to create a strong and durable seam. The study also found that the type of stitch used had a smaller impact on sewing thread consumption than fabric thickness. However, lockstitch was found to consume the most thread, followed by chain stitch, three-thread overedge stitch, and four-thread overedge stitch. Overall, the study concluded that fabric thickness is the most important factor affecting sewing thread consumption. The type of stitch used also has a significant impact, but to a lesser extent than fabric thickness (Yeşilpınar & Alkiraz, 2005).

Previous studies have shown that there are four tension peaks in the formation of a lockstitch 301 seam. These peaks occur when the needle descends and penetrates the fabric, when the thread is wrapped around the bobbin, when the needle rises and the thread is pulled taut, and when the thread is tightened to form the final stitch (Rengasamy & Samuel, 2011). Thread tension is an important parameter that affects the quantity of thread used for the seam in garment construction. If the thread tension is too high, the seam can be puckered or the thread can break. If the thread tension is too low, the seam can be loose and the thread can unravel. These can lead to more thread being used, as the seamstress will need to rethread the needle more often. By understanding how thread tension works and how it is affected by different factors, garment manufacturers can ensure that they are using the correct thread tension to produce high-quality seams with the optimum thread consumption.

2.2 Graphs, Tables And Formulas

The accurate measurement of thread lengths in stitches is achieved through careful unravelling under proper tension, following standardized methods like the French standard NF G07 101 (Jaouadi, et al., 2006). While this approach provides accurate results, it requires multiple stitching attempts to optimize sewing parameters for enhanced precision. This process is time-consuming and incurs testing costs for materials, equipment, and skilled labour. An experienced operator is essential to obtain precise yarn length values without subjecting them to excessive tension or distortions.

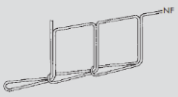
Due to the limitations of physically measuring thread consumption, researchers have explored alternative prediction methods for rapid and practical solutions. Various techniques such as value prediction charts, mathematical formulas, thread length ratios, predictive algorithms based on historical data, learning algorithms, and software solutions have been used to forecast thread consumption (Jaouadi, et al., 2006). However, when comparing the prediction results of these methods for the same stitch length and input parameters in a consistent stitch setup, significant variations in predictions have been observed.

Originally garment manufacturers often used a variety of graphs, tables and formulas to estimate thread consumption as shown in the Figure 2.1. (American & Efrid Inc, 2007; Amaan Group, 2010). However these graphs, tables and formulas are based on various assumptions and trial and error methods, providing less flexibility when used with varying fabric thickness and stitch densities, hence the accuracy of the predicted values was questionable. An alternate technique to calculate the thread consumption involved consumption ratios, which determine the thread amount relative to the stitch's geometry. Initially limited to one stitch density value, these ratios were customized for different stitches by thread suppliers, allowing accurate calculation of thread usage considering various stitch densities and fabric thicknesses (Amaan Group, 2010; American & Efrid Inc, 2007).

Leading thread suppliers are currently utilizing software packages to enhance the accuracy of calculating thread consumption (Abeysooriya & Wickramasinghe, 2014). These packages employ various formulas and ratios and can compute thread usage for diverse parameters like stitch lengths, stitch densities, and fabric thicknesses. They include definitions for the majority of stitch classes commonly used in the apparel industry, incorporating details like seam widths and appropriate stitch types.Unlike previous methods, these software solutions compute thread consumption separately for needle threads, bobbin threads, and looper threads. While some

software solutions account for thread properties like ticket number in their consumption calculations, many do not address essential physical properties associated with stitch formation.

AMANN sewing thread requirement tables

Stitch Type		Seam Construction ISO 4915 DIN 61400	Seam Appearance Top / Bottom	Seam Width mm	Stitches (per cm)	Thread Required per 1m of seam	%
Single-thread chainstitch	101			–	2	NF: 3,80 m	100 %
Single-thread blindstitch	103			–	2	NF: 4,50 m	100 %
Single-thread blindstitch	105			–	2	NF: 4,50 m	100 %
Lockstitch (Hand stitch type)	209			–	4	NF: 1,40 m	100 %

NF = Needle thread · GF = Bobbin/looper thread · LF = Cover thread

Remember to allow extra thread for beginning and end of seam

AVERAGE THREAD CONSUMPTION TOTALS BY GARMENT

The following is a list of sewn products and thread consumption totals based on thread consumption reports conducted by our Technical Service Department. These thread consumption figures include a 10% waste factor and are based on a typical garment construction.

Product Sewn	Total Yds/Garment	Product Sewn	Total Yds/Garment
Men's		**Boy's**	
Slack	250	Jeans	168
Jean	223	Pants	183
Jean Short	160	Jacket	175
Work Pants	200	Dress Shirt	94
Suit Coat	114	Knit Shirt	46
Dress Shirt – long sleeve	106	Baseball Cap	44
Work Shirt	171		
Knit Polo Shirt	135		
Fleece Sweat Shirt	262		
Tee Shirt	64		
Tank Top	38		
Knit Brief	52		
Women's		**Girl's**	
Lined Coat	246	Blouse	73
Blazer	153	Dress	118
Dress	141	Swim Suit	65
Skirt	192		
Blouse	122		
Pants	162		
Jeans	250		
Shorts	151		
Robe	300		
Night Gown	135		
Panties	62		
Bra	63		

Figure 2.1 AMANN and A&E sewing thread requirement tables

Coats, a company specializing in threads, introduced a software called SEAMWORKS, which calculates the amount of sewing thread used (Coats Digital, 2023). This program factors in various parameters, including stitch types and color groups, to determine the number of bobbins required for the thread consumption. Furthermore, SEAMWORKS provides a result report that includes the total cost for the used sewing thread amount.

2.3 Mathematical And Geometrical Models

To address the accuracy and flexibility issues in graphs, tables and ratios, researchers have suggested several theories and methods to precisely estimate thread consumption. These approaches usually entail developing mathematical and geometrical models to predict thread consumption based on an investigation of the variables that affect thread consumption, such as properties of thread and fabric and stitching parameters (Jaouadi, et al., 2006).

Some researches had the base of considering the geometric shape of the stitch type to predict thread consumption. One such study was conducted for the stitch type 301 lockstitch (Rasheed, et al., 2014). The mean absolute error between the actual and predicted thread consumption was calculated at 3% and the R^2 value of 0.97 showed that the actual consumption values were in good agreement with predicted consumption (Rasheed, et al., 2014).

Geometric models were introduced to predict thread consumption of different lockstitch shapes of lockstitch class 300, by studying the geometric shapes and relationships (Sarah, et al., 2020). As evidenced by the results, the models' accuracy was appreciably high with R^2 ranging from 93.91% to 99.1%. Moreover, stitch width, stitch density and the distance between two needles were discovered to be the most crucial factors influencing thread consumption, while yarn count and fabric thickness contributed less (Sarah, et al., 2020). Another geometrical model was developed for predicting thread consumption of stitch class 406 (Rehman, et al., 2021). Stitch class 400 use more thread than those of stitch class 300 but less amount of thread than those of stitch class 500. A total of 18 samples were sewed with various fabric types, SPIs (stitches per inch) and material thicknesses in order to validate the model. The model was found to have an accuracy of more than 97.18% in predicting sewing thread consumption (Rehman, et al., 2021).

Most existing geometrical models for predicting thread consumption are founded on rectangular profiles (Ghosh & Chavhan, 2014; Jaouadi, et al., 2006). The proposed geometrical model by (Rasheed, et al., 2014) and (Sarah, et al., 2020) also follows the rectangular profile, while considering interlacement point space and fabric diameter thickness. However, these models exhibit notable prediction errors likely due to the reliance on the rectangular seam shape and unrealistic assumptions like yarn cylindrical nature and fabric incompressibility.

In contrast, a more recent model proposed by (Chavan, et al., 2019) based on a realistic elliptical profile demonstrated comparatively lower error rates. A geometrical model for lockstitch seam 301 based on an elliptical profile was proposed to predict thread consumption irrespective of fabric type (Chavan, et al., 2019). This model also acknowledges the impact of seam

compression on the initial fabric assembly thickness, which in turn influences thread consumption. In comparison to other recent rectangular profile-based models, the model was tested across a variety of fabric types, stitch densities, and the number of fabric piles and proven to be more accurate and have the ability to generalize with less error (Chavan, et al., 2019). Three geometrical models namely elliptical, racetrack and circular were put out in a study to predict the consumption of thread in lockstitch seams (Chauhan & Ghosh, 2021). By comparing actual and predicted consumptions for samples sewn at various feed rates to validate the elliptical and racetrack models, the racetrack model was proven to be the most precise and adaptable. Also, it was discovered that the circular model worked well for samples with equal stitch spacing and fabric assembly thickness. The presented models provided a considerably lower error% than previous models (Chauhan & Ghosh, 2021).

A regression model was proposed to predict sewing thread consumption that included thread tension constraint and was evaluated on the lockstitch 301 and chain stitch 401 (Abeysooriya & Wickramasinghe, 2014). The study shows that thread tension plays a crucial role in determining thread consumption for both stitches. The combined impacts of thread tension, fabric thickness and stitch density in the case of chain stitch 401 determine precise sewing thread consumption taking the stitch's properties into account. The lockstitch 301 demonstrates a combined effect of yarn count and thread tension on the consumption of thread. The proposed regression model is expected to be a superior technique to calculate thread consumption of the 2 stitch types considered, as the inclusion of the thread tension parameter represented a reduction in error percentages (Abeysooriya & Wickramasinghe, 2014).

The focus of another study of (Khedher & Jaouachi, 2015) was on modelling sewing thread consumption in the context of manufacturing classic jeans. To precisely determine the amount of thread used in different types of stitching, they considered the contribution of waste factors. Several factors impact thread usage in sewn garments, including seam length, stitch density, seam types, and material thickness. However, these factors can vary depending on the specific garment style, leading to differing thread consumption in products like jeans, shirts, and jackets. This implies that thread consumption isn't a standard measure for various sewn product categories. The results of the linear regression method suggested a strong correlation between waste factors and experimental sewing thread consumption. This discovery reinforces the significance of waste factors as a crucial parameter when studying thread consumption in seamed denim jeans. Figure 2.2 shows the linear regression analysis indicating a very high R-squared (R^2) value of 99.8%, suggesting that waste factors explain 99.8% of the variation in thread consumption (Khedher & Jaouachi, 2015).

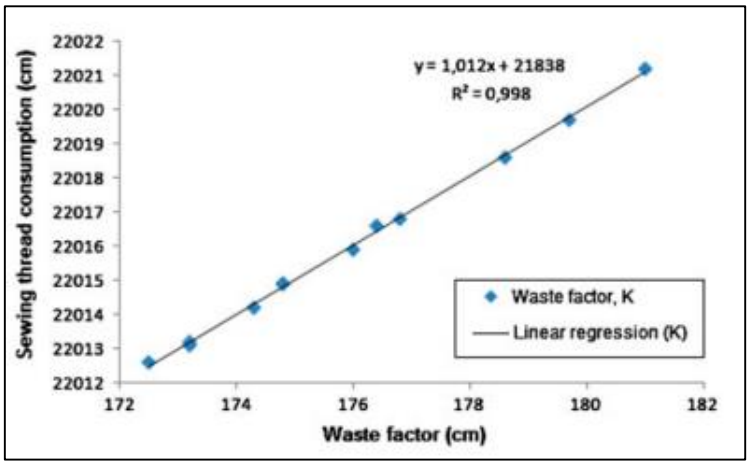


Figure 2.2 Relationship between the waste factor and experimental consumption values

Using stitch density, fabric assembly thickness and thread elongation another regression model has been created to predict the consumption of both types of sewing thread (cotton and polyester-cotton core spun threads) in chain-stitch stitch type (Sharma , et al., 2017). The proposed model can be utilized to predict thread consumption with an R^2 of 0.956 and an ability to explain 95.6% of the variation. This eliminated the need to employ several models for various thread types.

Another study by (Mariem, et al., 2020) introduced a multilinear regression model to predict thread consumption for women's undergarments. While regression models tend to be more accurate, they are highly specific to certain material types and sewing conditions.

2.4 Machine Learning, Neural Network Models And Metaheuristic Optimization Models

Another study was carried out involving three different analytical models namely, the theoretical model, regression model and artificial neural network model to predict the amount of sewing thread required for different stitch types, stitch densities and thread types (Jaouadi, et al., 2006). As per the study, the model using the artificial neural network model was the most reliable at estimating the sewing thread consumption in a garment. Parameters such as fabric thickness, stitch density (number of stitches per centimetre) and sewing thread count served as the basis for the development of the model. The study employed 4 different stitch types namely lockstitch, two-thread chain stitch, three-thread chain stitch and safety stitch along with 4 different stitch densities ranging from 3 to 6 stitches per centimetre and 2 distinct cotton sewing threads. The study highlighted that the artificial neural network model performs best when it comes to estimating thread consumption with an accuracy of at least 95%. It further emphasized that as the artificial neural network approach is capable of retraining when new data from

another area of the training data domain become available, it is anticipated to perform substantially better (Jaouadi, et al., 2006).

It was observed that for defining, analysing constraints, making predictions and modelling both complex and non-linear issues, a fuzzy theory-based analytical model allows significantly more flexibility than traditional techniques such as regression, neural network, mathematical and subjective (Jaouachi & Khedher, 2013). The results using the fuzzy theory show that the sewing thread consumption values are influenced by the thread composition, which is represented by the number of twisted yarns and the kind of fibre, the size of the needle and the fabric weight. The fuzzy forecast model was produced following the correlations between input and output parameters that were determined using Mandani's min-max inference. A better understanding of the influence that input parameters have on the sewing thread consumption of jean pants was provided by the generated fuzzy rules (Jaouachi & Khedher, 2013).

Another study by Jaouachi & Khedher focused on developing a fast and precise way to predict the amount of sewing thread needed for making a garment involving two modelling approaches: linear regression and artificial neural networks. The effectiveness of each model was assessed by comparing the predicted thread consumption to actual values obtained from unstitching garments, using the R-squared coefficient (Jaouachi & Khedher , 2015). Various statistical measures were analysed to enhance the linear regression model. However, it was found that the linear regression method falls short in accurately predicting the thread usage variation for jean trousers. On the other hand, the neural network technique proved to be effective in explaining consumption variation in our experimental setup. The trained neural network's performance was assessed by evaluating errors in training, validation, and test sets. Their findings indicate that the neural model achieved a regression coefficient of 0.973, closely approaching 1 as in the above (Jaouachi & Khedher , 2015). This high coefficient reflects the minimal mean error between actual and predicted consumptions. This suggests that the developed neural network model is effective in accurately predicting the thread consumption for jean trousers. Figure 2.3 illustrates the accuracy of values obtained from the validation sample in testing the neural network model. These results align well with the conclusions of Jaouadi et al.'s study (Jaouadi, et al., 2006).

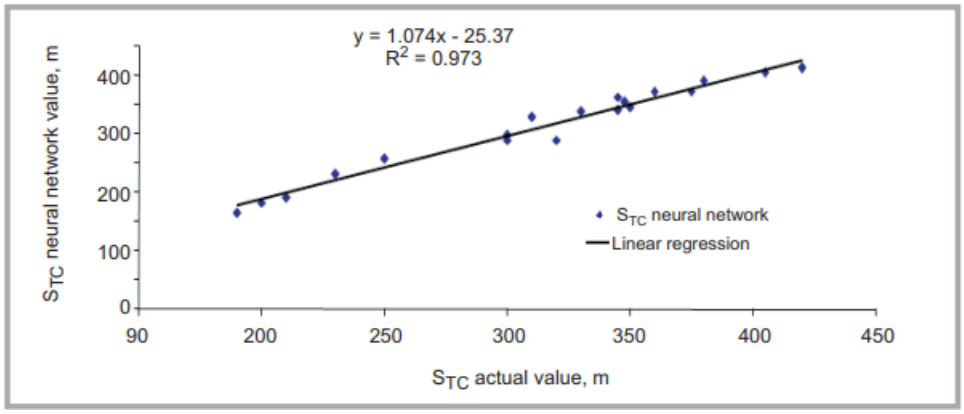


Figure 2.3 Post-training analysis of neural network results

The study of (Chavhan, et al., 2021) introduced regression models and an Artificial Neural Network (ANN) model to predict how much thread is needed for the assembly of multiple layers of different fabrics using lock stitch 301. These predictions were compared with existing methods for the same type of stitch. Among the available methods, empirical predictions show a high 59% error, while geometric models exhibit errors ranging from 8% to 26% for different multilayer compositions. Regression models show more variation in prediction accuracy. The proposed quadratic regression model has an absolute percent error of 4.3%, and the ANN network 1 displays an overall absolute error of 2.1%, though it reaches 9.5% during testing (Chavhan, et al., 2021). Both the quadratic regression and neural network models offer more accurate predictions than other methods. They are also versatile, able to predict thread consumption for different types of fabric assemblies. In industries using various fabrics and assembly types, ANN models can be customized and trained accordingly. With more data and continuous training, their accuracy can improve further.

A different approach to reducing the amount of sewing thread required to make a pair of jeans was followed by employing three different metaheuristic techniques namely particular swarm optimization (PSO), ant colony optimization (ACO) and genetic algorithm (GA) (Jaouachi & Khedher, 2022). Results revealed that PSO and ACO methods were more precise than experimental methods and the GA method, and had the lowest thread consumption with the optimal input parameter combinations to sew jeans.

Nonetheless, the accuracy of estimating thread consumption has increased with the introduction of mathematical, geometrical and machine learning-based models. Regression models, artificial neural networks and fuzzy theory-based models have all been employed in the studies to predict thread consumption and the results demonstrate that machine learning-based models outperform conventional approaches (Jaouachi & Khedher, 2013).

Therefore, this study aims to analyse new variables that affect sewing thread consumption in a garment that is being manufactured at the author's organization, in addition to those already studied, and subsequently develop a novel initiative to be employed at the organization, to accurately predict thread consumption in a garment with the aid of machine learning techniques.

2.5 Chapter Summary

The literature review explores various methodologies employed in estimating sewing thread consumption in garment manufacturing. The apparel industry's growing focus on cost efficiency to maintain competitiveness underscores the significance of accurately calculating material usage. The factors influencing thread consumption, such as garment type, size, design, fabric, stitch length, stitch density, and stitch type, are discussed comprehensively.

Researchers have traditionally used graphs, tables, formulas, and consumption ratios, but these methods often lack flexibility and accuracy across different fabric types and stitch densities. Alternative approaches, including mathematical and geometrical models, have been proposed to enhance precision. Geometrical models based on realistic shapes, such as elliptical and racetrack profiles, have shown improved accuracy compared to rectangular profiles.

Despite the accuracy improvements achieved with mathematical and geometrical models, the literature emphasizes that machine learning-based models, particularly artificial neural networks, provide superior accuracy, reaching at least 95%, by considering parameters like fabric thickness, stitch density, and sewing thread count. Therefore the study will delve into the evolution of predictive models, highlighting the transition from traditional techniques to machine learning and artificial neural network models.

CHAPTER 3

METHODOLOGY AND THEORY

This chapter serves as the guiding framework for the systematic exploration and understanding of the factors influencing sewing thread consumption in garment manufacturing. The first section will cover the steps of the research framework. The following sections will concentrate on the theoretical aspects elucidating the statistical and machine learning methods employed. The chapter consists with a few metrics that can be utilized to evaluate the performance of the models built. Finally the chapter comes to an end with a list of architectural decisions that have been made to design an effective and reliable solution, while addressing the complexities inherent in the apparel manufacturing process. Through a systematic approach, the Methodology and Theory chapter sets the stage for the subsequent analyses, evaluations, and the ultimate deployment of the predictive model in a real-world operational context.

3.1 Methodology

In this section, the intricate steps undertaken in the study were presented to develop a predictive model for sewing thread consumption in the garment manufacturing domain, focusing on TKT 120 and TKT 160 thread types (Figure 3.1).

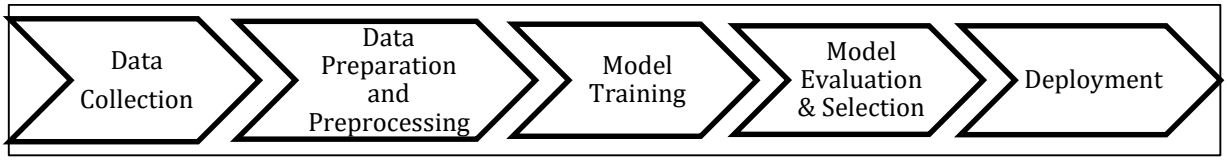


Figure 3.1 Proposed Methodology

3.1.1 Data Collection

The data for this study was collected from Emjay International (Pvt) Ltd., a BOI-approved apparel manufacturing company which engages mostly in B2B (business-to-business) business serving the best high street retailers both in Europe and in the U.S.A (Emjay International and Penguin Sportswear, 2020). The organization has been recording thread consumption for each style they receive from the customers for many years, thus the amount of data accessible for this study is extensive.

3.1.1.1 Data Sources

The data is extracted from two primary sources. They are data from the ERP (Enterprise Resource Planning) System and style-wise thread consumption worksheets.

The ERP system is a consolidated database that houses all the data required for business operations. This database will give users access to a multitude of data on the manufacturing process from costing and ordering to production-based operations such as cutting, sewing, packing and shipment. The relevant style details which are yet to be produced, was extracted from the ERP system and the parameter data that is used to derive the thread consumption was extracted from the thread consumption worksheets handled by the Product Development team

The style-wise thread consumption worksheet format used in the organization was originally developed by A&E (American & Efird Inc.) (Inc, 2007) and altered by the Product Development team in order to align with the organization's processes and policies. These worksheets include specific parameters that are used to calculate the thread consumption of a garment such as the thread types and stitch types that should be used in each operation in a garment (Operation Breakdown – OB), fabric thickness and the machine types.

3.1.1.2 Data Collection Process

The data collection process is a critical phase in this study aimed at predicting thread consumption in garment manufacturing. It involves gathering, acquiring, and preparing the necessary data to develop accurate predictive models. Figure 3.2 outlines the key aspects of the data collection strategy.

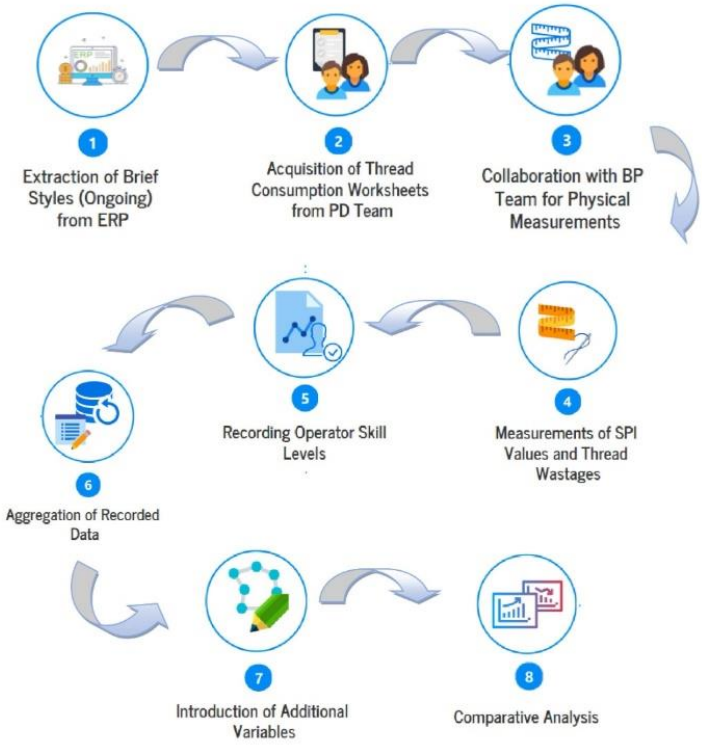


Figure 3.2 Data Collection Process Flow Diagram

Initially, the full brief styles which are to be produced was extracted from the ERP based on their production start dates and subsequently, acquired the corresponding thread consumption worksheets from the Product Development (PD) team (Figure 3.3). A collaborative effort was established with the Business Process (BP) team to conduct physical measurements for several critical parameters in the garment production process, specifically, focused on measuring and tracking wastages for each type of thread used in each operation. Moreover, the study had a human dimension which recognized the pivotal role of machine operators in shaping the outcome of the production processes. Operator skill levels were registered in a dedicated database named as Machine Operator (MO) Grading Report. This information was linked to the specific operators and the operations considered for the completion of full brief styles of medium-sized garments.

3.1.2 Data Preparation And Preprocessing

The data that was physically recorded during the measurements was systematically aggregated into a format consistent with the thread consumption worksheet received from the Product Development team (Figure 3.3). Additionally, certain variables were introduced to effectively represent the data obtained through physical measurements. This harmonization of data structures was undertaken with the specific purpose of facilitating a comprehensive comparative analysis between the estimated thread consumption, as determined by the Product Development team, and the actual thread consumption as recorded through physical measurements (Figure 3.4). This comparative assessment is integral to this study, enabling a thorough evaluation of the accuracy and alignment of estimations with real-world data. Afterwards, the collected data were pre-processed and prepared for the analysis.

Oper #	Name of Operation	Machine Type	Number of M/C Aallo.	ISO Stitch	Rows of Stitch	SPI	Seam Length cm	Seam Thickness mm	Needle Thread cm	Needle Thread	Bobbin Thread cm	Bobbin Thread Tex / Type / Color	Looper Thread cm	Looper Thread Tex / Type / Color	Cover Thread cm	Cover Thread Tex / Type / Color	Total cm/ Oper.
1	Front & Back Gusset Attach	OL	1	514 (4mm) 2 Ndl Overed	1	14	18.00	1.5	95.5	TKT 120 COL 1			236.8	TKT 160 COL 1			332.3
2	Elastic Attach To Leg	DEVISC	2	406 3.2mm Btm Cover	2	16	79.00	1.2	833.5	TKT 120 COL 1			1704.1	TKT 160 COL 1			2537.7
3	Elastic Attach To Back Waist	DEVISC	1	406 3.2mm Btm Cover	1	16	58.00	1.2	306.0	TKT 120 COL 1			625.6	TKT 160 COL 1			931.5
4	Side Seam With Label	OL	3	514 (5mm) 2 Ndl Overed	2	14	32.00	1.0	269.1	TKT 120 COL 1			937.0	TKT 160 COL 1			1206.1
5	Side Tack X 4	SNLS	3	301 Lockstitch	4	14	8.00	3.2	88.4	TKT 120 COL 1	88.4	TKT 120 COL 1					176.9
6	Cut Lace	MANUAL		Stitch #?													
7	Lace Attach To Front Waist	FL	1	406 3.2mm Btm Cover	1	16	65.00	1.2	342.9	TKT 120 COL 1			701.1	TKT 160 COL 1			1044.0
8	Front & Back Gusset Attach	OL	1	514 (4mm) 2 Ndl Overed	1	14	24.00	1.5	127.4	TKT 120 COL 1			315.7	TKT 160 COL 1			443.1

Figure 3.3 PD's Thread Consumption Worksheet

Name of Operation	Operator Skill	Machine Type	Number of M/C Aallo.	ISO Stitch	Rows of Stitch	SPI	Seam Length cm	Seam Thickness mm	TKT 120 COL 1	TKT 120 COL 1 Wastage cm	Total TKT 120 COL 1	TKT 160 COL 1	TKT 160 COL 1 Wastage cm	Total TKT 160 COL 1	TKT 120 Wastage cm (PD)	TKT 160 Wastage cm (PD)
Gusset Attach - Front	A+	OL	1	514 (4mm) 2 Ndl Overedge	1	14	18	1.5	95.53	1.91	97.44	236.79	4.74	241.52	3.8	9.5
Elastic Attach To Leg	A+	FL DEVISOR	2	406 3.2mm Btm Cover	2	16	79	1.2	833.54	16.67	850.21	1,704.12	34.08	1,738.21	33.3	68.2
Elastic Attach To Back Waist	A	FL DEVISOR	1	406 3.2mm Btm Cover	1	16	58	1.2	305.98	6.12	312.10	625.56	12.51	638.08	12.2	25.0
Side Seam With Label	C	OL	3	514 (5mm) 2 Ndl Overedge	2	14	32	1	269.10	5.38	274.48	937.03	18.74	955.77	10.8	37.5
Side Tack X 4	C	SNLS	3	301 Lockstitch	4	14	8	3.2	176.88	3.54	180.42	0	0	0	7.1	-
Cut Lace		MANUAL							-	-	-	-	-	-	-	-
Lace Attach To Front Waist	B	FL	1	406 3.2mm Btm Cover	1	16	65	1.2	342.91	6.86	349.77	701.06	14.02	715.09	13.7	28.0
Gusset Attach - Back	C	OL	1	514 (4mm) 2 Ndl Overedge	1	14	24	1.5	127.37	2.55	129.92	315.72	6.31	322.03	5.1	12.6

Figure 3.4 Harmonized Data Set

The Thread Consumption Dataset is a comprehensive collection of data meticulously gathered and curated to facilitate the proposed research and analysis. The descriptions and the data types of the 26 variables are listed in the below table (Table 3.1).

Table 3.1 Dataset Description

Attribute index	***Attribute name***	***Attribute description***	***Attribute type***
1	Style	The specific garment style or category to which each data point belongs.	Categorical
2	Buy	The specific number of the repeated order received from the customer.	Categorical
3	Name of Operation	Specific sewing or manufacturing operation performed on the garments during production.	Categorical
4	Operator Skill	The skill level of the machine operator involved in the manufacturing process.	Categorical
5	Machine Type	The type of sewing machine used for a particular operation	Categorical
6	Number of Machines Allocated	The quantity or count of sewing machines used for a specific operation.	Numerical
7	Stitch type	The type of stitch used in the operation, providing a standardized reference.	Categorical
8	Stitch ID	A unique identifier associated with each type of stitch	Categorical
9	Needle	A mathematical formula which helps to get the thread quantity per centimeter considering SPI and Seam thickness for the needle thread	Numerical
10	Bobbin	A mathematical formula which helps to get the thread quantity per centimeter considering SPI and Seam thickness for the bobbin thread	Numerical
11	Looper	A mathematical formula which helps to get the thread quantity per centimeter considering SPI and Seam thickness for the looper thread	Numerical

12	Rows of Stitch	The number of rows of stitches within a given operation	Numerical
13	SPI	The density of stitches per linear inch in the garment	Numerical
14	Seam Length cm	The length of the seam produced during a particular operation in centimeters	Numerical
15	Seam Thickness mm	The thickness of the seam or the fabric you are sewing together in milimeters	Numerical
16	Needle Thread cm	Needle thread consumption, measured in centimeters, reflects the quantity of thread used for the needle thread. (Seam Length * Number of Rows of Stitches * N)	Numerical
17	Needle Thread	Additional details about the needle thread, which may include information such as thread type, ticket number, and color	Categorical
18	Bobbin Thread cm	Bobbin thread consumption, in centimeters, reflects the amount of thread used for the bobbin threads. (Seam Length * Number of Rows of Stitches * B)	Numerical
19	Bobbin Thread Tex /Type/Color	Additional details on the Bobbin thread, which include information such as thread type, ticket number, and color	Categorical
20	Looper Thread cm	Looper thread consumption, in centimeters, reflects the amount of thread used for the looper threads	Numerical
21	Looper Thread Tex /Type/Color	Additional details on the Looper thread, which include information such as thread type, ticket number, and color (Seam Length * Number of Rows of Stitches * L)	Categorical
22	Estimated Wastage %	The estimated percentage of thread wastage during the production process	Numerical
23	TKT 120 COL 1	The amount of TKT 120 thread used during the operation, measured in centimeters. This includes both needle thread and bobbin thread.	Numerical
24	Total TKT 120 COL 1	The total consumption of TKT 120 thread, including both used and wasted thread, in centimeters	Numerical
25	TKT 160 COL 1	The amount of TKT 160 thread used during the operation, measured in centimeters.	Numerical
26	Total TKT 160 COL 1	The total consumption of TKT 160 thread, including both used and wasted thread, in centimeters	Numerical

The data preparation and preprocessing phase in the study is a critical step to ensure that the collected data is accurate, complete, and suitable for analysis. This phase involves several key processes as follows.

3.1.2.1 Data Cleaning And Handling Missing Data

One of the initial steps is to eliminate duplicates in the dataset to avoid redundancy.

Null values were present in the dataset which belonged to the “Cut Lace” Operation. Thus all the rows which had the "Cut Lace" operation were excluded from the dataset. This decision was made because this operation does not involve sewing and therefore does not require thread. Including it in the analysis would introduce noise and potentially skewed results.

For fields where no threads from the bobbin and loopers were utilized (specifically, in the "Bobbin" and "Looper" columns), zero-imputation approach was employed thus replacing missing values with zeros, indicating that no thread was used in those cases.

Symbols in the data, such as dashes, were transformed into numerical values (e.g., zeros). This transformation standardizes the data, making it suitable for mathematical operations and analysis.

Two columns, "Number of Machines Allocated." and "Stitch ID," were omitted from the analysis. This decision was made because these columns are irrelevant to the final output of the study and including them would not contribute to the research objectives.

3.1.2.2. Data Transformation

Data is transformed into the format required by machine learning algorithms. Data pivoting, mapping, encoding, normalization, and integration are all examples of this. The categorical variables such as Operation, Machine Type and Operator skill are transformed to numerical variables with the aid of one-hot encoding and label encoding techniques.

3.1.2.3 Feature Engineering

In this research, feature engineering plays a pivotal role in enhancing the dataset's relevance and effectiveness for predicting sewing thread consumption across seven distinct sewing operations involving two different thread types. To accommodate this complex task, new fields have been derived by leveraging existing ones (Figure 3.5).

Name of Operation	Operator Skill	Rows of Stitch	SPI	Seam Length cm	Seam Thickness mm	Needle Thread cm	Needle Thread	Bobbin Thread cm	Bobbin Thread Tex / Type / Color	Looper Thread cm	Looper Thread Tex / Type / Color	TKT 120 COL 1	TKT 120 COL 1 Wastage cm	Total TKT 120 COL 1	TKT 160 COL 1	TKT 160 COL 1 Wastage cm	Total TKT 160 COL 1
Gusset Attach - Front	A+	1	14	18	1.5	95.53	TKT 120	-		236.79	TKT 160	95.53	1.91	97.44	236.79	4.74	241.52
Elastic Attach To Leg	A+	2	16	79	1.2	833.54	TKT 120	-		1,704.12	TKT 160	833.54	16.67	850.21	1,704.12	34.08	1,738.21
Elastic Attach To Back Waist	A	1	16	58	1.2	305.98	TKT 120	-		625.56	TKT 160	305.98	6.12	312.10	625.56	12.51	638.08
Side Seam With Label	C	2	14	32	1	269.10	TKT 120	-		937.03	TKT 160	269.10	5.38	274.48	937.03	18.74	955.77
Side Tack X 4	C	4	14	8	3.2	88.44	TKT 120	88.44	TKT 120	-		176.88	3.54	180.42	0	0	0
Cut Lace						-		-		-		-	-	-	-	-	-
Lace Attach To Front Waist	B	1	16	65	1.2	342.91	TKT 120	-		701.06	TKT 160	342.91	6.86	349.77	701.06	14.02	715.09
Gusset Attach - Back	C	1	14	24	1.5	127.37	TKT 120	-		315.72	TKT 160	127.37	2.55	129.92	315.72	6.31	322.03

Figure 3.5 Feature Engineered Data set

Here's a detailed breakdown of the feature engineering process.

1. Thread Type Differentiation - The study involves two thread types, TKT 120 and TKT 160, utilized for needle thread, bobbin thread, and looper thread. To facilitate separate tracking and analysis of thread consumption for each thread type, distinct fields have been created for each operation.

2. TKT 120 Thread Fields - The TKT 120 COL 1 field represents the quantity of TKT 120 thread consumed during an operation, measured in centimeters. This is derived by summing the values of the existing variables Needle Thread cm and Bobbin Thread cm. Additionally, the 'Total TKT 120 COL 1' field has been generated, which accounts for wastage. This is achieved by adding the 'TKT 120 COL 1' and 'TKT 120 COL 1 Wastage cm' fields.

3. TKT 160 Thread Fields - The TKT 160 COL 1 field corresponds to the amount of TKT 160 thread utilized during an operation, also measured in centimeters. This is equivalent to the Looper Thread cm field. Similar to the TKT 120 thread, the Total 'TKT 160 COL 1' field has been formulated to incorporate wastage. It is calculated by adding the 'TKT 160 COL 1' and 'TKT 160 COL 1 Wastage cm' fields.

These derived fields allow for a granular analysis of thread consumption for each thread type across the sewing operations. By distinguishing between TKT 120 and TKT 160 threads the dataset becomes more informative and suitable for the predictive modeling of sewing thread consumption in a comprehensive and accurate manner.

After feature engineering was done few more columns such as "Bobbin" and "Looper Thread Tex / Type / Color" columns which did not add value to the study were omitted.

3.1.3 Model Training

As the next step, different machine learning techniques will be applied to the cleaned, preprocessed data and subjected to training. As in the literature review provided, different machine learning and neural network models will be trained to assure that the best machine learning model is adopted for predicting the actual thread consumption with the desired level of accuracy of 95% or above. (Jaouadi, et al., 2006; Jaouachi & Khedher, 2013).

The preprocessed thread consumption data will then be trained on below models and the performance measures were retrieved for both training and test data.

3.1.3.1 Multivariate Linear Regression

Multivariate Linear Regression is an extension of simple linear regression, capable of modeling relationships between multiple independent variables and a dependent variable. It seeks to find the best-fitting linear equation to predict the outcome based on the input features.

3.1.3.2 Random Forest

Random Forest is an ensemble learning method that constructs a multitude of decision trees during training and outputs the average prediction (regression) or the most frequent prediction (classification) of the individual trees. It excels in handling complex relationships in data.

3.1.3.3 Gradient Boost

Gradient Boosting is an ensemble learning technique that builds a series of weak models sequentially, with each model correcting the errors of the previous one. It aims to minimize a loss function, resulting in a strong predictive model.

3.1.3.4 XGBoost

XGBoost (Extreme Gradient Boosting) is a highly efficient and scalable implementation of gradient boosting. It incorporates regularization techniques and parallel processing, making it a popular choice for various machine learning tasks, known for its speed and performance.

3.1.3.5 Multilayer Perceptron (Neural Network)

Multilayer Perceptron, a type of artificial neural network, consists of multiple layers of interconnected nodes (neurons). It is a versatile and powerful model capable of learning complex patterns and relationships in data, often employed in tasks such as image recognition and natural language processing.

3.1.4 Model Evaluation and Selection

After the models have been trained, the accuracy and effectiveness of the models will be evaluated using a test data set. The author will use various measures such as mean absolute error (MAE) and root mean square error (RMSE) between the predicted and actual thread consumption (Jaouadi, et al., 2006; Abher, et al., 2014). These evaluation metrics are generally acknowledged and employed in the machine-learning community for model evaluations. MAE and RMSE will measure the magnitude of errors in the model's predictions. MAE provides the average error magnitude, whereas RMSE gives greater weight to large errors. Using these measures will enable benchmarking and comparison between different models. After the evaluation of the models, the best model will be selected based on the highest accuracy, thus, ensuring the selected model is robust and have the ability to generalize to new data which is essential for the model's performance in the longer run.

3.1.5 Deployment

As the final step, the evaluated model will be introduced to the business's operational environment. The author will collaborate with the Product Development team to assure that the model is incorporated into the thread consumption prediction process and used efficiently to output the optimum result.

3.2 Theory

This section outlines the theories of statistical and machine learning methods used in carrying out the research.

3.2.1 Statistical Models

In the pursuit of understanding and predicting sewing thread consumption, the study employs various statistical models to capture the intricate relationships between key variables. In this section, three fundamental types of statistical models were introduced, each serving a unique purpose in uncovering patterns and dependencies within the dataset.

3.2.1.1 Simple Linear Regression Model

The relationship between a dependent variable Y and a single independent variable X is quantified in simple linear regression (Olive, 2017). The simplest form of relationship between two variables is a straight line and the relationship between Y and X should be defined in the form of a simple linear regression model as

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Here, β_0 and β_1 are called model regression coefficients or parameters, and ε is a random disturbance or error.

It is assumed that this linear equation provides an acceptable approximation to the true relationship between Y and X within the considered range of observations and ε measures the discrepancy in that approximation. In particular ε is assumed to contain no systematic information for determining Y that has not already captured in X.

Standard regression assumptions

- The response Y and the explanatory variables are linearly related.
- ε is a normally distributed random variable with mean zero $(i.e.\ E[\varepsilon_i] = 0)$ and constant variance σ^2 $(i.e.\ Var[\varepsilon_i] = \sigma^2)$ for all $i = 1, 2, 3, \dots, n$.
- ε_i and ε_j are uncorrelated so that the covariance between ε_i and ε_j is zero $\left(i.e.\ Cov[\varepsilon_i\ ,\ \varepsilon_j\ ] = 0\right)$ for all $ij; i \neq j$.

3.2.1.2 Multiple Linear Regression

Similarly in Multiple regression, several predictors, independent or explanatory variables are used to model a single response (dependent) variable (Olive, 2017). The the model can be given in matrix form as ,

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

Same assumptions of the linear regression models apply here.

$$\boldsymbol{E(\varepsilon) = 0}$$

Here, $E(\varepsilon_i) = 0$ for each $i = 1, 2, \dots, n$ can be given in a matrix form as above.

$$\boldsymbol{V(\varepsilon) = \sigma^2 I_n} = \begin{pmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}$$

Similarly to above $V(\varepsilon_i) = \sigma^2$ for each $i = 1, 2, \dots, n$ can also be given as a diagonal matrix with σ^2 in the diagonal and zeros in off-diagonal elements as covariance between error terms are assumed to be zero in the model.

$$\boldsymbol{\varepsilon \sim N(0, \sigma^2 I_n)}$$

Overall, the errors follow a multivariate normal distribution with mean vector 0 and variance-covariance matrix $\boldsymbol{\sigma^2 I_n}$. (Here I_n is the $(n \times n)$ identity matrix).

3.2.1.3 Multivariate Linear Regression

Multivariate linear regression is a technique of modeling multiple responses with a single set of predictors (Olive, 2017). The model have the following form,

$$Y_{n \times p} = X_{n \times (k+1)} \beta_{(k+1) \times p} + \varepsilon$$

The model is constructed based on following assumptions,

- The variables of interest are linearly related with the response variables.
- There are no outliers.
- Error (ε) is normally distributed with zero mean and constant variance.
- Two or more variables are not substantially correlated with each other which presence of the correlation refers to the multicollinearity.

3.2.2 Machine Learning Models

Artificial Intelligence is a continuously growing field in computer science and can be referred to as the ability of computers or machine in carrying out tasks which are associated with intelligent beings (Copeland, 2023). The entire space of artificial intelligence consists of many components. Machine and deep learning are two aspects within this space as depicted below.

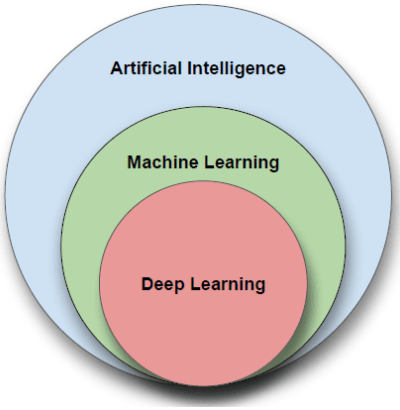


Figure 3.6 Artificial Intelligence

(Source: (Abhishek, 2022))

Machine Learning (ML) learns from data and experience to improve the performance of tasks and decision making using statistical methods whereas Deep Learning (DL) uses complex neural network structures. According to the Figure 3.6, deep learning can be considered a part of machine learning and is involved with the use of artificial neural networks. These have outperformed ML models in the instances where the accuracy is considered important over interpretability because neural networks are harder to interpret.

3.2.2.1 Ensemble Models

Ensemble modeling is the process of developing diverse amount of models to predict an outcome, either by using many different modeling algorithms or using different training datasets. The final prediction for the unseen data is then produced by the ensemble model by combining the predictions of all the base models. The goal of employing ensemble models is to lower the prediction's generalization error. Using the ensemble approach reduces the model's prediction error provided that the base models are independent and diversified. The method bases its predictions on the wisdom of the multitude. The ensemble model functions and behaves as a single model even when it has several foundation models. Random forest and boosting algorithms are such examples for ensemble modeling.

Random Forest

As mentioned above, random forest is an ensemble machine learning model that can be applied to both regression and classification problems. It is an advanced version of decision trees where multiple random trees are modelled to construct the final ensemble model. The final prediction is the mean or the majority vote of these individual predictions depending on whether it is a regression or a classification problem respectively.

Leo Breiman introduced random forest as a solution for the increase in variance caused by correlated trees when attempting bagging (Breiman, 2001)The solution was to take samples of the variables in each split which would result in reduced correlations because it reduces the chances of two trees being similar in each step.

Mutual Information Criteria

Mutual information (MI) is a nonlinear measure to quantify the statistical dependency between two or more variables. MI takes zero, if the two variables are independent and if they are dependent, it takes a positive value reflecting the strength of the relationship. (Barraza, et al., 2018)state that it is more suitable than the cross correlation to quantify a dependency between variables that differ from linear relationships.

Feature selection using random forest

The methodology of this study was followed up by a feature selection since random forest provides the facility to plot the feature importance. Feature selection involves the selection of the most relevant features in a dataset used in predictions. It is important since it provides a platform where the model's complexity could be reduced for the same level accuracy in predictions. The interpretability and reduced overfitting are two more benefits of conducting feature selection.

The importance of a variable is defined by how much the feature reduces the impurity. The impurity in regression is measured by the variance. The more significant a feature is to the predictive model, the higher the reduction in impurity it causes. The importance of a feature can be assessed by averaging the decrease in impurity for that feature across all trees.

Hyperparameter Tuning

The parameters that were tuned in this study are as follows, the rest remained as their default values.

- max_depth – The longest path between the leaf node and its root node.
- n_estimators – The number of trees in the forest.
- min_samples_split – The minimum number of samples required to split an internal node.
- min_samples_leaf – The minimum number of samples required to be at a leaf node.

Gradient Boosting

Gradient Boosting is a powerful boosting algorithm that combines several weak learners into strong learners, in which each new model is trained to minimize the loss function such as mean squared error or cross-entropy of the previous model using gradient descent (Natekin & Knoll, 2013). Each time around, the algorithm calculates the gradient of the loss function in relation to the current ensemble's predictions, trains a new weak model to minimize this gradient, and repeats the process. The procedure is then continued until a stopping criteria is satisfied after the predictions of the new model have been added to the ensemble.

The algorithm tries to learn the function $f(x)$ that maps the input features X to the target variable Y. $f(x)$ is the sum of the boosted trees. The loss function $L(f)$ is the difference between the actual and the predicted variables.

$$L(f) = \sum_{i=1}^{N} L(y_i, f(x_i))$$

The loss function is minimized with respect to f as,

$$\hat{f}_0(x) = \underset{f}{\operatorname{argmin}}\, L(f) = \underset{f}{\operatorname{argmin}} \sum_{i=1}^{N} L(y_i, f(x_i))$$

If the gradient boosting algorithm is in M stages then to improve the f_m the algorithm can add some new estimator as h_m having $1 \leq m \leq M$

$$\hat{y}_i = F_{m+1}(x_i) = F_m(x_i) + h_m(x_i)$$

For M stage gradient boosting, the steepest descent finds $h_m = -\rho_m g_m$ where ρ_m is constant and known as step length and g_m is the gradient of loss function $L(f)$,

$$g_{im} = \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x_i)=f_{m-1}(x_i)}$$

Similarly, the gradient for M trees is,

$$f_m(x_i) = f_{m-1}(x_i) + \left(\underset{h_m \in H}{\operatorname{argmin}} \left[\sum_{i=1}^{N} L\big(y_i, f_{m-1}(x_i) + h_m(x_i)\big)\right]\right)(x)$$

The solution will be,

$$f_m = f_{m-1} - \rho_m g_m$$

XGBoost

XGBoosting is a more regularized version of gradient boosting. It uses advanced regularization ($L1\ and\ L2$), which improves the model generalization capabilities. It is also known to deliver high performance as compared to gradient boosting and it trains faster and can be parallelized across clusters.

3.2.2.2 Multilayer Perceptron (Neural Network)

An Artificial Neural Network (ANN) is a biologically inspired sub-field of artificial intelligence which attempts to mimic the human brain with less sophistication. It is a layered, feedforward and completely connected network of neurons. A fully connected multi-layer neural network is called a Multilayer Perceptron (MLP) (Noriega, 2005).

The most basic MLP consists of three layers of which the initial layer is the input layer, and the final layer is the output layer. The layer in the middle is called the hidden layer. The decision of how many hidden layers to use in a model depends on the intuition of the architect of the neural network. Higher the number of hidden layers, the more the complexity of the network which results in slowing down the process. The number of nodes in the input layer depends on the number of features in the dataset and number of output nodes depends in the classification or prediction problem at hand.

Each of these nodes are connected to every node in the adjacent layer. The connected neuron has an associated weight and a threshold that determines its activation or deactivation. Neuron activates itself, if the threshold is achieved. Framework of an Artificial Neural Network is shown below in the Figure 3.7.

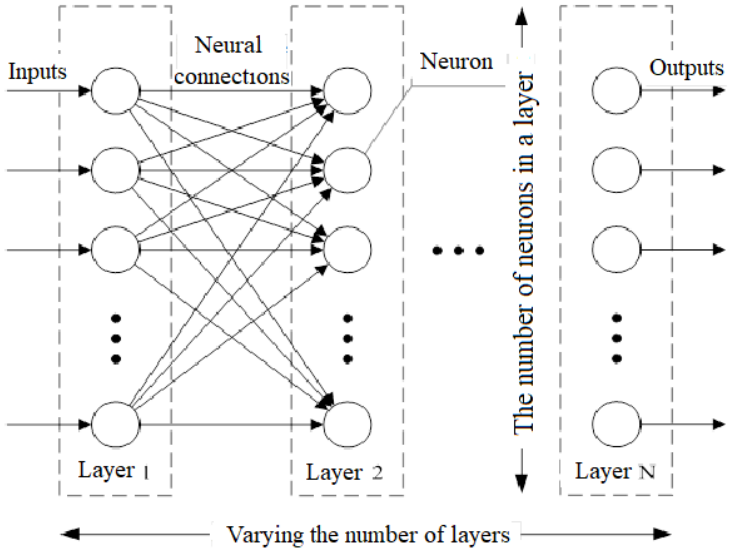


Figure 3.7 Framework of an Artificial Neural Network

(Source: (Vasiliev, et al., 2019))

3.2.3 Model Comparison

The following error measures were used to compare the predictive power of the models predicting the closing prices and the returns of the individual US stocks employed in the study.

1. RMSE – Root mean squared error

   This is calculated by taking the square root of the average squared difference between the actual and the predicted value.

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

   - y_i – Actual values
   - $\hat{y}_i$ – Predicted values
   - n – Number of observations

   RMSE is relatively easy to compute but it can be affected by outliers or skewed data.

2. MAE – Mean Absolute Error

   MAE is a measure of the average size of the error in a collection of predictions, without considering the error direction. It is computed as below,

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

3. MAPE – Mean Absolute Percentage Error

   MAPE is another commonly used performance metric in regression problems. It is computed as below,

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

Compared to RMSE, MAE and MAPE are comparatively less sensitive towards outliers or extreme points.

3.3 Research/Solution Design

Several architectural decisions have been made to design an effective and reliable solution. These decisions are grounded in the rationale of achieving accurate predictions while addressing the complexities inherent in the apparel manufacturing process. The decisions and the rationale behind them are listed in the Table 3.2.

Table 3.2 Architectural Decisions & Rationale

Decision	***Description***	***Rationale***
Data Integration	The solution integrates data from various sources, including ERP systems, Product Development teams for thread consumption worksheets, and Business Process teams for physical measurements of SPI and wastages and recording operator skill data.	Thread consumption prediction relies on a multitude of parameters such as garment styles, machine types, operator skills, and thread types. Integrating data from diverse sources ensures a comprehensive understanding of the production process.
Feature Engineering	The solution involves extensive feature engineering to create new variables representing thread consumption for different thread types (e.g., TKT 120 and TKT 160) in each operation.	By deriving new features from existing data, the model can capture the specific consumption patterns of different thread types, enhancing prediction accuracy.
Multivariate Linear Regression	Multivariate Linear Regression will be employed as one of the primary modeling techniques.	Multivariate Linear Regression extends linear regression to predict multiple dependent variables (thread consumptions of TKT 120 and TKT 160) simultaneously.
Ensemble Models	Ensemble modeling techniques like Stacking and Bagging will be adopted to combine predictions from multiple models.	Ensemble methods are known for their ability to improve prediction accuracy by leveraging the strengths of different models. In the apparel industry, where thread consumption patterns can be complex, ensembles offer a robust solution.
Operator Skill Incorporation	Operator skill levels are integrated into the analysis, acknowledging their impact on production outcomes.	Operator skill is a critical factor influencing thread consumption. Incorporating this data allows for a more nuanced understanding of the production process.

3.4 Chapter Summary

The Methodology and Theory chapter lays the groundwork for accurate sewing thread consumption prediction in garment manufacturing. The chapter starts with a framework that explains the steps carried out in the study. The subsiding sections presents the theoretical and methodological background of the analysis. Leveraging a robust data integration approach from ERP systems, Product Development and Business Process teams, the study employs extensive feature engineering and adopts modeling techniques, including Multivariate Linear Regression, Ensemble Models and Neural Network models. Theory section brings attention to the theoretical aspects on the statistical measures and machine learning methods employed during the study.

CHAPTER 4

EXPLORATORY DATA ANALYSIS

In order to gain a better understanding of the data distribution and characteristics and to find trends and correlations in the data, exploratory data analysis (EDA) will be carried out.

In the context of thread consumption analysis, the primary focus is on assessing whether the estimates for thread usage, specifically for two thread types, TKT 120 and TKT 160, obtained from the Product Development team might have been overestimated. Thus a hypothesis testing was conducted.

In the conducted hypothesis test, the author aimed to assess whether there is a statistically significant difference between the estimated thread consumptions retrieved from the PD's thread consumption worksheets (Estimated TKT 120 COL 1 & Estimated TKT 160 COL 1) and the actual thread consumptions (TKT 120 COL 1 & TKT 160 COL 1). Formulated hypotheses are as follows.

- For TKT 120: Null Hypothesis (H_0): Estimated TKT 120 COL 1 <= TKT 120 COL 1
  Alternative Hypothesis (H_α): Estimated TKT 120 COL 1>TKT 120 COL 1
- For TKT 160: Null Hypothesis (H_0): Estimated TKT 160 COL 1 <= TKT 160 COL 1
  Alternative Hypothesis (H_α): Estimated TKT 160 COL 1 > TKT 160 COL 1

The statistical analysis employed a one-tailed t-test to compare the means of the Estimated thread consumptions for TKT 120 and TKT 160 and the actual thread consumptions from the dataset. Following the t-test and evaluation of the p-value, for both cases TKT 120 and TKT 160, the null hypothesis is rejected. This means that there is robust statistical evidence indicating that the estimated thread consumption values (Estimated TKT 120 COL 1 and Estimated TKT 160 COL 1) are significantly greater than the actual thread consumption values (TKT 120 COL 1 and TKT 160 COL 1) for the operations or scenarios encompassed within the dataset.

4.1 Buffer Wastage Inbuilt By The Product Development Vs Measured Wastage

The comparison between the inbuilt wastages provided by the product development team and the measured wastages reveals a notable discrepancy (Figure 4.1). Across various operations and for both TKT 120 and TKT 160 thread types, the inbuilt wastages consistently appear significantly higher than the measured wastages. This observation suggests potential inaccuracies or overestimations in the wastage estimations provided by the product development team. Such discrepancies could lead to inefficiencies in resource allocation and

may impact the overall production process, including material management and cost estimation. Therefore it becomes imperative to reassess and refine the wastage estimations to align them more closely with the actual measured wastages, enabling better-informed decision-making and resource utilization throughout the production lifecycle.

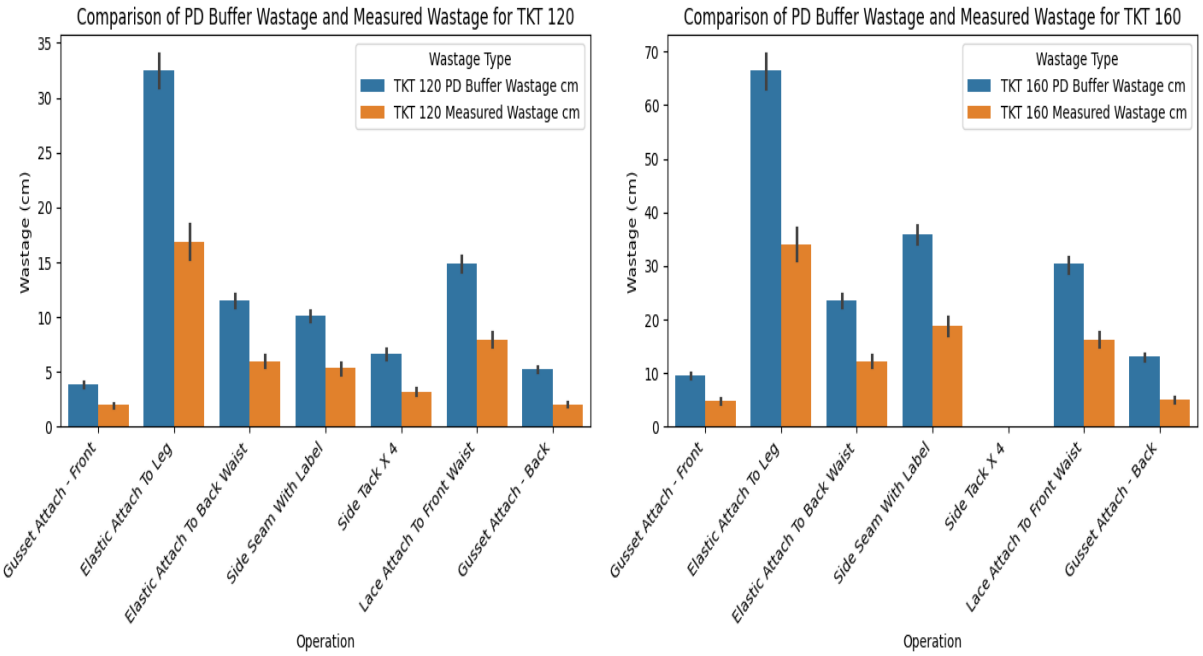


Figure 4.1 Buffer Wastage inbuilt by the Product Development vs Measured Wastage

4.2 Influence Of Spi On The Consumed Thread Amount Behavior

SPI (Stitches per inch) represents the density of stitches within a given length, and its influence on thread consumption is substantial. A higher SPI generally results in increased thread usage (Figure 4.2), as more stitches are packed into a fixed seam length. This is attributed to the fact that a greater number of stitches requires more thread to traverse the same distance.

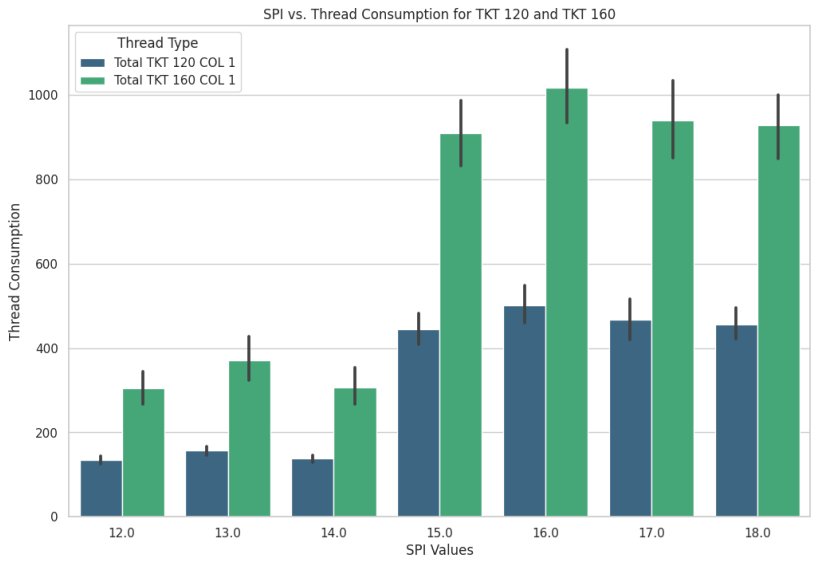


Figure 4.2 Influence of SPI on the thread consumption

4.3 Influence Of Seam Length On The Consumed Thread Amount Behavior

Seam length refers to the measurement of the stitched line or joint formed by sewing two pieces of fabric together, that is the distance along the sewn line and is typically measured in linear units such as centimeters or inches. A longer Seam Length typically requires a greater amount of thread to secure the fabric, resulting in increased thread consumption as shown in Figure 4.3. The clusters observed in the Scatter Plot (Figure 4.3) depicting the relationship between Seam Length and thread consumption convey meaningful insights regarding the thread utilized across specific seam lengths within the seven distinct operations. Each cluster encapsulates a set of data points that share a similar pattern of thread consumption concerning the corresponding seam length.

Figure 4.3 Influence of Seam Length on the thread consumption

4.4 Influence Of Seam Thickness On The Consumed Thread Amount Behavior

Seam thickness refers to the combined measurement of the layers of fabric joined together by a seam, representing a pivotal dimension in the creation of stitched structures within a garment. This parameter is particularly significant as it directly affects the mechanical properties of the seam, including its strength, durability, and overall integrity. The thickness of the seam is closely associated with the complexity of the sewing process, as thicker seams often demand a greater quantity of thread to ensure secure stitching. Additionally, variations in seam thickness can influence the tension and stress distribution along the seam, further impacting the overall thread consumption pattern. The correlation matrices obtained for each operation in Figure 4.4 delves into the relationship between seam thickness and the thread consumption of both TKT

120 and TKT 160. Positive correlations signify that an increase in seam thickness is associated with a higher thread consumption.

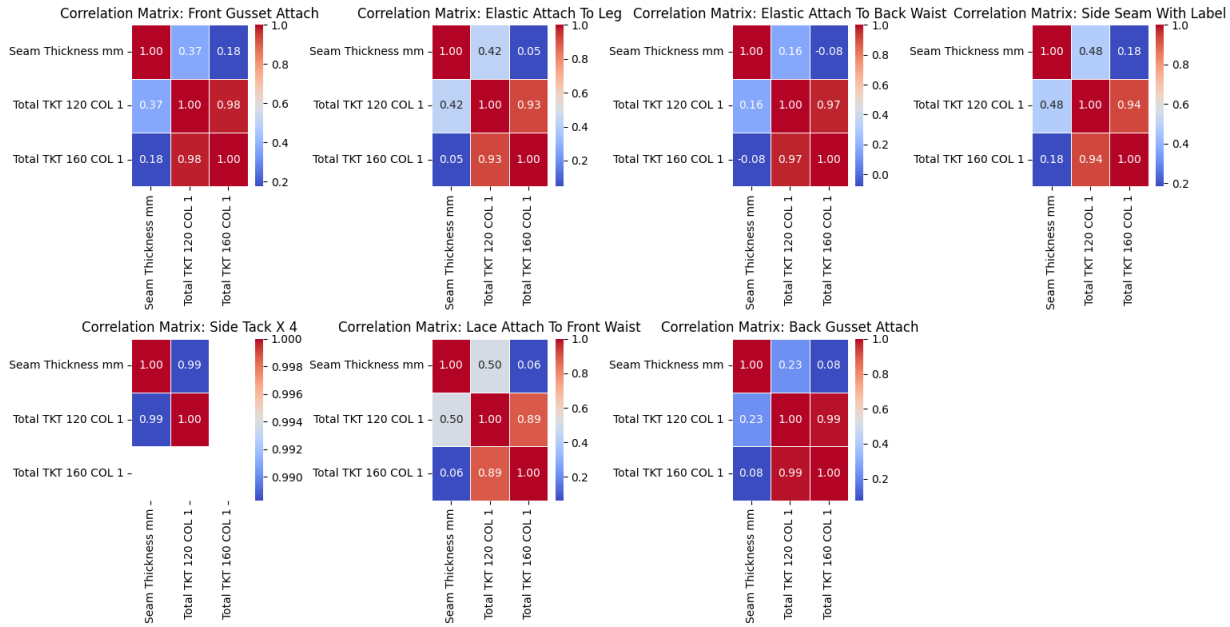


Figure 4.4 Correlations between Seam Thickness and thread consumption

4.5 Influence Of Wastage On The Consumed Thread Amount Behavior

Wastage in the thread consumption for a garment refers to the proportion of thread that is expected to be wasted during the manufacturing process. It represents the amount of thread that is used but does not contribute to the final stitched product. Positive correlations in the Figure 4.5 indicate that as wastages increase, there is a corresponding elevation in thread consumption.

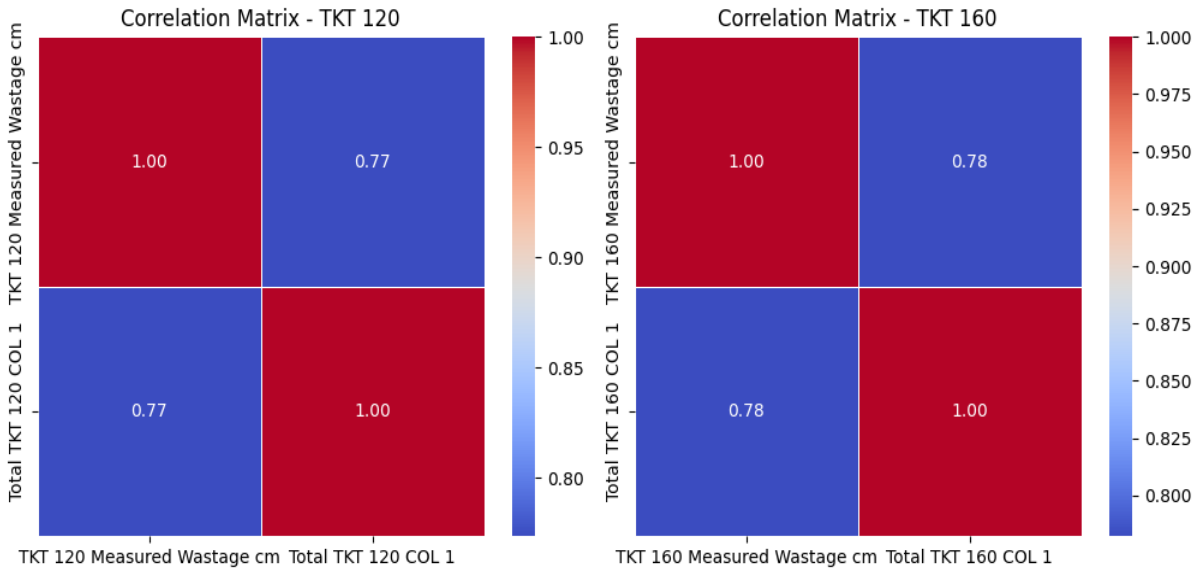


Figure 4.5 Correlation between Wastage and thread consumption

4.6 Influence Of Operator Skill On The Consumed Thread Amount Behavior

When it comes to thread consumption, operator skill describes the ability and knowledge of the people in charge of sewing duties in the apparel industry. The amount of thread used in various stitching operations can be greatly influenced by the operators' skill level. Higher grades, such an A+,A and B are often reserved for skilled operators who have a deeper knowledge on the sewing machinery, procedures, and specifics of the sewing operations they do. Their proficiency allows them to maximize the use of thread, reducing waste and guaranteeing effective sewing as evidenced from the Figure 4.6.

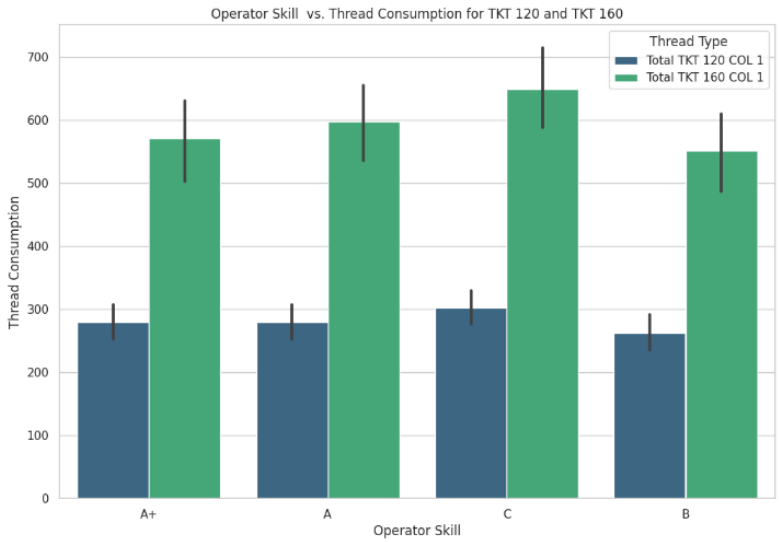


Figure 4.6 Influence of Operator Skill on the thread consumption

The Figure 4.6 suggests a positive relationship between operator skill levels and thread consumption, where higher skill grades (A+, A and B) tend to be associated with more efficient and optimized thread usage, while lower skill grade C exhibits a tendency towards increased thread consumption.

4.7 Chapter Summary

This chapter explores exploratory data analysis (EDA) to scrutinize thread consumption estimates, particularly for TKT 120 and TKT 160, provided by the Product Development team. The hypothesis test revealed a significant overestimation in the thread consumption. A comparison between inbuilt wastages and measured wastages underscored a substantial disparity, attributing potential inaccuracies. While factors like SPI, seam length, seam thickness, and operator skills remained stable, the critical factor elevating thread consumption is the addition of buffer wastage. This analysis emphasizes the need to refine estimations of buffer wastages and highlights buffer wastages as a key factor significantly impacting overall thread consumption in garment production.

CHAPTER 5

EVALUATION AND RESULTS

As the next step, different machine learning techniques were applied to the cleaned, preprocessed data and subjected to training. In light of the substantial buffer wastage inherent in the product development team's estimations, there arises a critical necessity to forecast wastages based on actual measured wastage data. Subsequently, leveraging these predicted thread wastages becomes integral in accurately forecasting thread consumptions. In the context of predicting the wastages and consumptions of two thread types (TKT 120 and TKT 160) for the same garment, several machine learning models could be employed. These models are designed to handle multi-thread type predictions simultaneously. This chapter provides a comprehensive view on the forecasting models carried out and the results obtained during the research.

5.1 Multivariate Linear Regression

Multivariate Linear Regression is an extension of the simple linear regression model, designed to handle multiple dependent variables concurrently. In this research, it can be applied to predict the consumptions of both TKT 120 and TKT 160 threads simultaneously. This model assumes a linear relationship between various independent variables (such as stitch length, stitch density, seam length, etc.) and the thread consumptions. It allows for the modeling of how changes in these independent variables affect the thread consumptions of both thread types.

Before the multivariate linear regression model was fit, the preprocessing steps highlighted in the section 3.1.2, were carried out. For TKT 120, all the operations were considered for the training where as for TKT 160, the operation which had zero consumption from TKT 160 was excluded from training. The buffer wastage amount was predicted before arriving at the final consumptions for both thread types. Moreover a feature selection was carried out referring to the mutual information scores plot obtained (Figure 5.1).

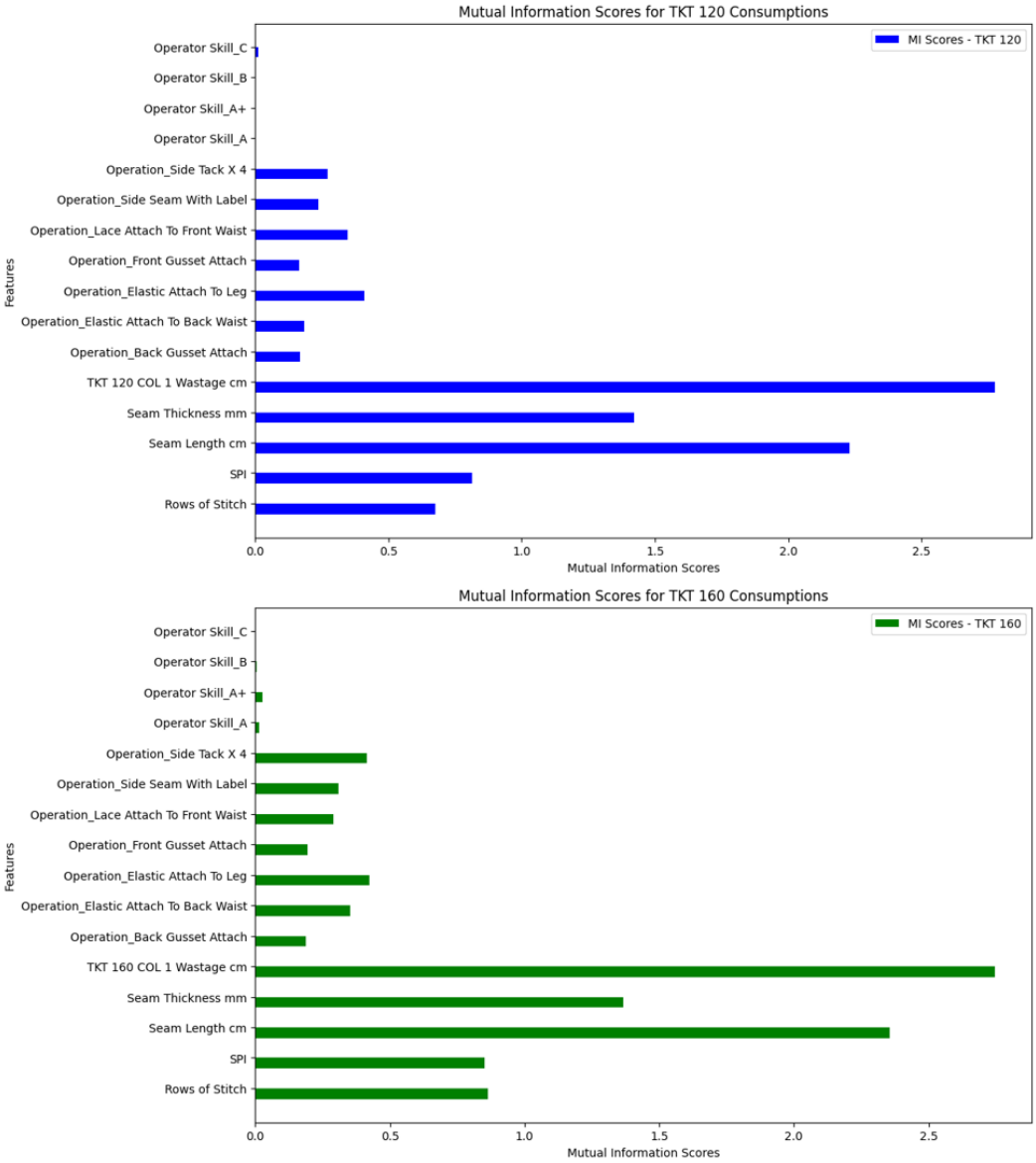


Figure 5.1 Mutual Information Scores Plot for Linear Regression models

The Table 5.1 and Table 5.2 show the results obtained for the RMSE, MAE and R-squared matrices for training and test data sets before and after the modification and feature selection for the wastage model and the consumption model respectively.

Table 5.1 Evaluation matrices of Linear Regression Models - Wastage Model

Metrics for Wastage Model	***Before Feature Selection (with all the variables)***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	3.99	4.61	9.94	7.89	3.99	4.59	9.94	7.90
MAE	2.65	2.99	6.96	5.81	2.67	2.97	6.97	5.79
MAPE	57.97%	61.90%	60.28%	57.43%	58.33%	61.00%	60.28%	57.27%
R-Squared	0.6175	0.4078	0.5183	0.5648	0.6160	0.4124	0.5174	0.5633

Table 5.2 Evaluation matrices of Linear Regression Models - Consumption Model

Metrics for Consumption Model	***Before Feature Selection (with all the variables)***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	15.61	16.29	29.57	25.42	15.61	16.28	29.55	25.39
MAE	10.25	9.89	22.74	19.98	10.26	9.89	22.72	19.96
MAPE	4.79%	4.18%	4.12%	4.07%	4.78%	4.17%	4.12%	4.06%
R-Squared	0.9951	0.9946	0.9960	0.9966	0.9951	0.9946	0.9960	0.9966

As per the tables 5.1 and 5.2, the model performance measured from each evaluation matric before and after the feature selection has not had a significant change in all evaluation matric values. However the error values obtained for the test data set have decreased by a slight amount in both wastage and consumption models. MAPE values for the wastage model are relatively high, suggesting a notable level of discrepancy between the predicted and actual values. Conversely, the consumption model exhibits lower MAPE values, indicating a closer alignment between predictions and actual observations. However, it's important to note that low MAPE values alone don't necessarily signify a superior model.

The actual and the predicted values for thread consumptions of the thread types TKT 120 and TKT 160 obtained from the linear regression models were plotted and represented in the below Figure 5.2. Slightly high deviations in the actual thread consumptions values were observed against the predicted values for thread type TKT 120 when compared to actual vs predicted values deviations of thread type TKT 160.The deviations were observed for high consumption values.

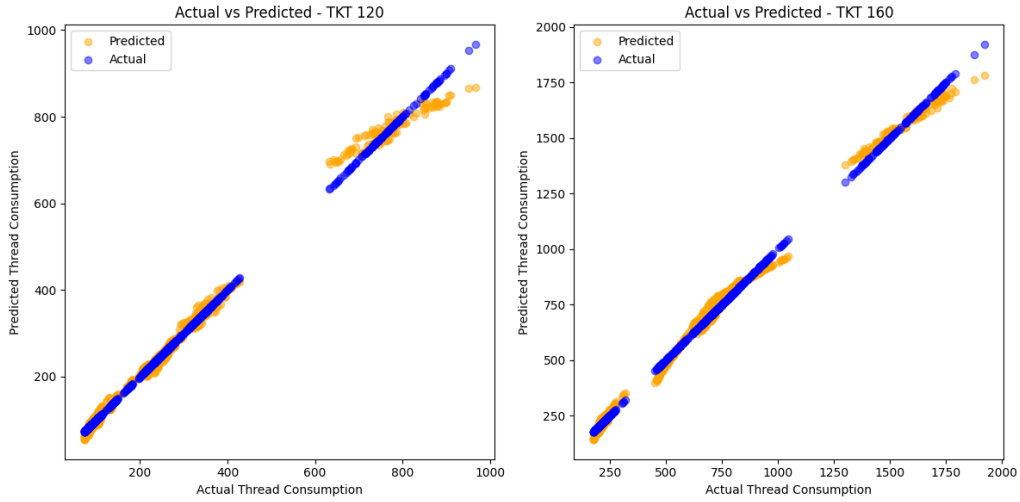


Figure 5.2 Actual vs Predicted values of Linear Regression model

5.2 Ensemble Models

This approach involves combining predictions from multiple individual models to create a more accurate and robust prediction. By leveraging the strengths of different models, ensemble methods like Stacking and Bagging can significantly enhance the accuracy and reliability of thread consumption forecasts.

5.2.1 Random Forest

Random Forest is an ensemble learning method that can be adapted for regression tasks involving multiple dependent variables. It combines the predictions of multiple decision trees, making it robust, less prone to overfitting and accurate for predicting thread consumptions for both TKT 120 and TKT 160. Random forest regressor was used to train the thread consumption data for both types of thread TKT 120 and TKT 160. Hyperparameters that were mentioned in the section 3.2.2 under random forest were tuned using grid search approach for every model built. Further a mutual information scores plot (MI Scores plot) was generated to identify the features that are closely connected to the target variables.

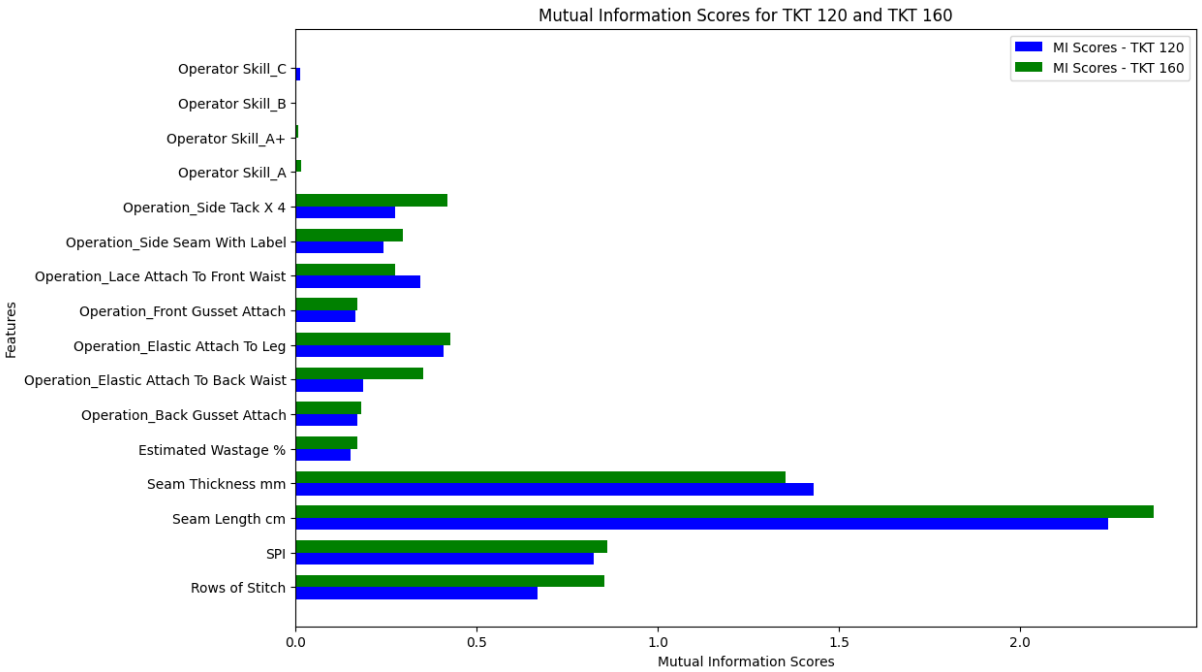


Figure 5.3 Mutual Information Scores Plot for Random Forest model

First the models were trained considering all the variables then followed by a feature selection with the aid of the MI Scores plot in Figure 5.3. As indicated by the MI scores plot in Figure 5.3, the thread consumptions of both thread types are less likely to be dependent on the Operator Skill levels. Therefore the model was trained excluding the operator skill feature. The results

obtained for the matrices taken into account before and after the feature selection are depicted in the table 5.3 and 5.4 below.

Table 5.3 Evaluation matrices of Random Forest model - Wastage Model

Metrics for Wastage Model	***Before Feature Selection (with all the variables)***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	3.50	5.91	8.94	10.50	2.53	5.97	6.57	10.29
MAE	2.35	3.65	6.12	7.20	1.64	3.70	4.27	7.36
MAPE	53.28%	75.29%	54.89%	70.29%	36.64%	75.55%	37.43%	70.01%
R-Squared	0.7050	0.0250	0.6098	0.2301	0.8451	0.0050	0.7892	0.2592

Table 5.4 Evaluation matrices of Random Forest model - Consumption Model

Metric for Consumption Model	***Before Feature Selection (with all the variables)***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	3.66	7.31	9.14	10.58	3.08	7.91	7.27	11.64
MAE	2.43	4.22	6.27	7.31	2.01	4.51	4.80	8.13
MAPE	0.88%	1.35%	0.90%	1.12%	0.70%	1.43%	0.68%	1.24%
R-Squared	0.9997	0.9989	0.9996	0.9994	0.9998	0.9987	0.9997	0.9992

As per the Table 5.3 and 5.4, the evaluation matrices of random forest algorithm shows significant reduction in error values than that of the multivariate linear regression model for both wastage and consumption models. However same behavior as the multivariate regression model was observed in the MAPE values generated for wastage prediction from the random forest model. The lowest MAPE value 37% was obtained only after feature selection for training datasets of both thread types. Moreover the evaluation matrices for training dataset are lower than that of the test dataset indicating very high performance on the training set, but there is a noticeable increase in error values that is a drop in performance on the test set. This suggests potential overfitting, as the models may have memorized the training data rather than generalizing well to new, unseen data. After the feature selection, where operator skill is removed, the results of the evaluation matrices of the test data set has increased more, indicating that the performance of the model to the unseen data after feature selection has degraded more.

However both RMSE and MAPE values of the training data set have decreased after the feature selection indicating the performance of the model on training data have increased.

The actual and the predicted values for thread consumptions of the thread types TKT 120 and TKT 160 obtained from the random forest models were also plotted and represented in the below Figure 5.4. In both the consumption values belonging to the two thread types, no significant deviations were identified indicating that there is a possibility of overfitting nature.

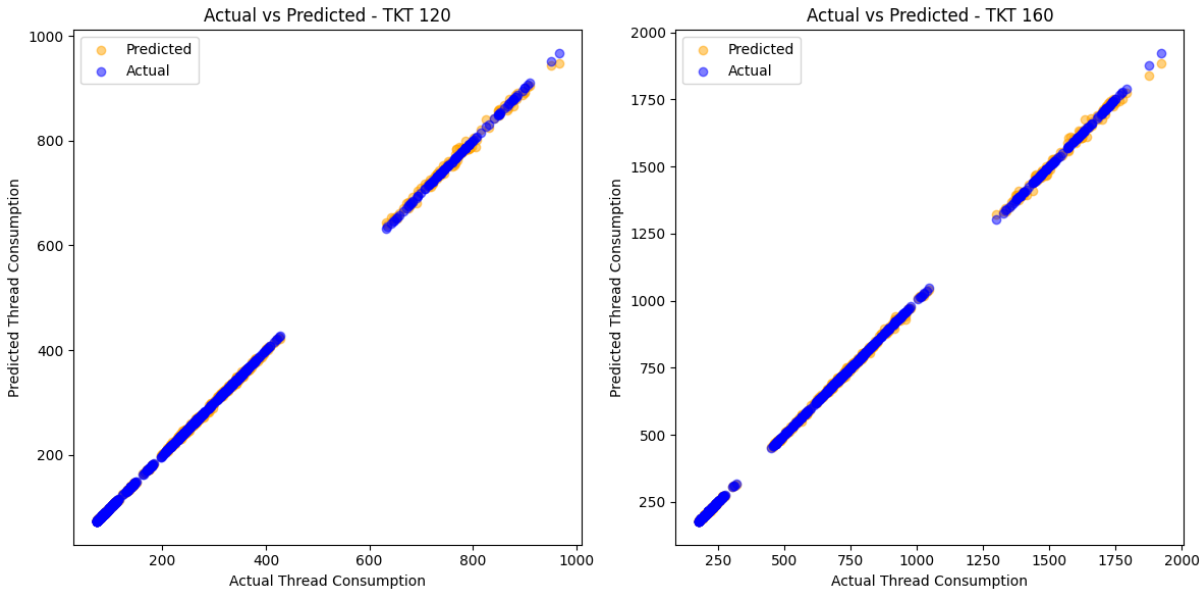


Figure 5.4 Actual vs Predicted values of Random Forest model

5.2.2 Gradient Boosting

Gradient Boosting is a powerful ensemble learning technique that enhances model performance by sequentially fitting weak learners, usually decision trees, to correct errors from the previous models. It iteratively minimizes a loss function, placing greater emphasis on instances with prediction inaccuracies. The wastages and the thread consumptions of both thread types TKT 120 and TKT 160 were trained using the GradientBoostingRegressor with all the features and reduced number of features. The results of the performance measures pre and post feature selection are represented in the Table 5.5 and 5.6.

Table 5.5 Evaluation matrices of Gradient Boosting model - Wastage Model

Metrics for Wastage Model	***Before Feature Selection (with all the variables)***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	3.71	5.25	9.27	9.33	3.53	5.45	8.45	9.15
MAE	2.49	3.21	6.48	6.35	2.4	3.24	6.06	6.33
MAPE	56.55%	64.43%	57.90%	61.17%	55.09%	64.85%	55.66%	59.97%
R-Squared	0.6688	0.2307	0.5803	0.391	0.7002	0.1717	0.6513	0.4149

Table 5.6 Evaluation matrices of Gradient Boosting model - Consumption Model

Metric for Consumption Model	***Before Feature Selection (with all the variables)***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	5.68	7.33	12.63	13.41	5.66	7.29	12.90	13.91
MAE	3.99	4.67	8.99	9.72	4.00	4.68	9.16	10.03
MAPE	1.68%	1.66%	1.41%	1.76%	1.68%	1.65%	1.39%	1.71%
R-Squared	0.9994	0.9989	0.9992	0.9990	0.9993	0.9989	0.9992	0.9989

As per the Table 5.5 and 5.6, the model performance measured from each evaluation matric before and after the feature selection has not had a significant change in all the evaluation matric values.

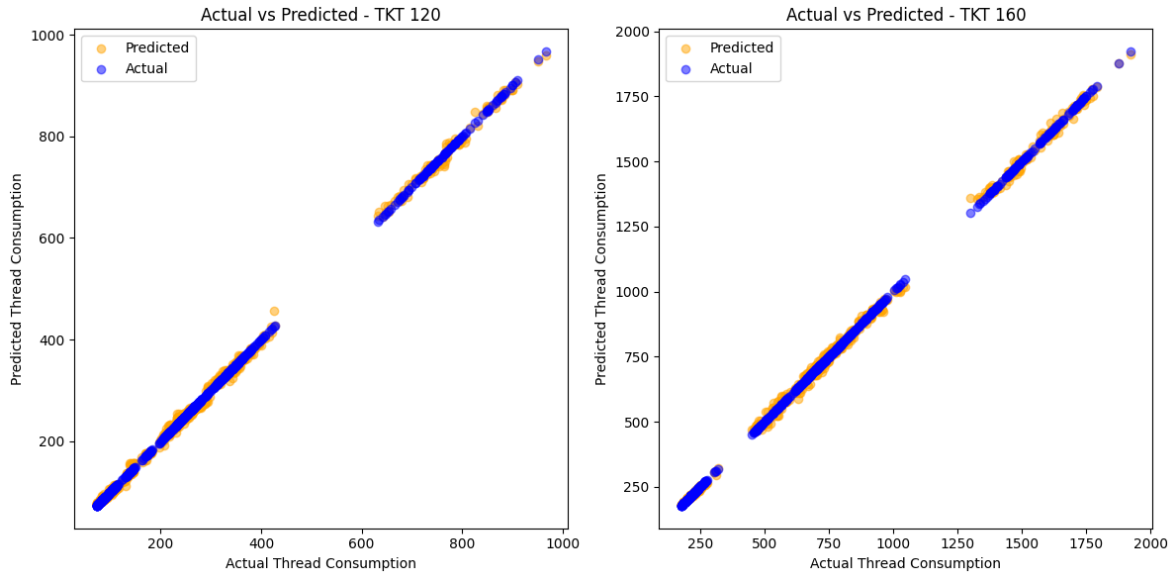


Figure 5.5 Actual vs Predicted values of Gradient Boosting model

Slight deviations were observed in the actual vs predicted values in the Gradient Boosting model for both thread types as demonstrated in the Figure 5.5.

5.2.3 XGBoost

XGBoost is renowned for its efficiency and effectiveness and excels in handling complex relationships within datasets. Its ensemble learning framework combines the strengths of multiple decision trees, making it adept at capturing intricate patterns and nuances in the data. XGBRegressor was used to train the thread consumption data of both the thread types TKT 120 and TKT 160, followed by a feature selection via a featue importance plot for both the thread types (Figure 5.6)

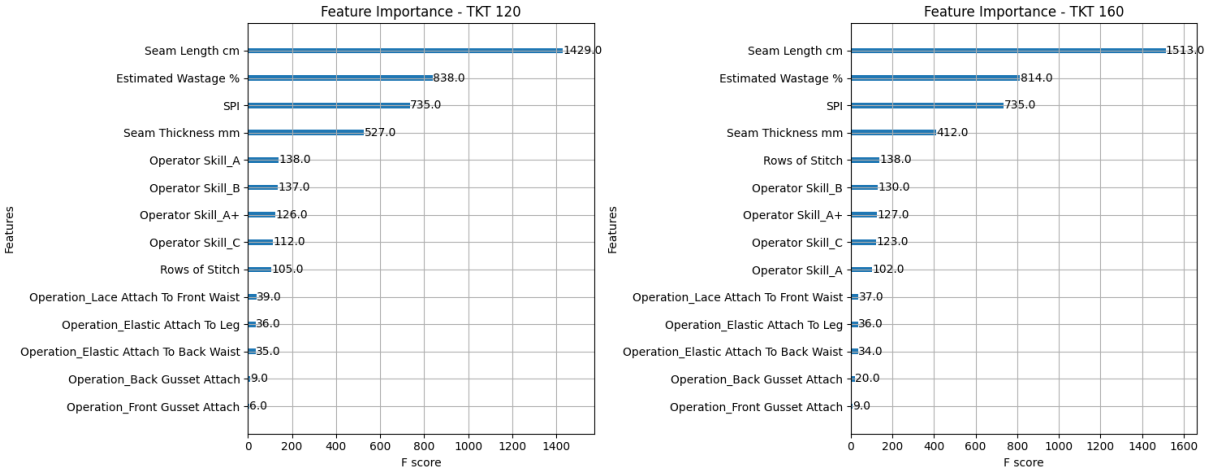


Figure 5.6 Feature Importance Plot for XGBoost model

The outcomes derived from the matrices considered both prior to and subsequent to the implementation of feature selection are elegantly presented in Table 5.5 below.

Table 5.7 Evaluation matrices of XGBoost model - Wastage Model

Metrics for Wastage Model	***Before Feature Selection (with all the variables)***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	3.46	6.5	8.89	11.49	2.23	6.82	5.77	11.68
MAE	2.31	3.91	6.06	7.67	1.26	4.13	3.2	8.33
MAPE	52.52%	78.69%	54.23%	74.05%	32.22%	79.68%	30.55%	75.32%
R-Squared	0.7105	0.7098	0.6143	0.6032	0.8802	0.8652	0.8372	0.8067

Table 5.8 Evaluation matrices of XGBoost model - Consumption Model

Metric for Consumption Model	***Before Feature Selection (with all the variables)***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	2.26	7.03	5.82	11.81	2.21	6.50	5.46	10.12
MAE	1.32	4.23	3.30	8.48	1.24	3.91	3.26	6.89
MAPE	0.84%	1.30%	0.87%	1.17%	0.54%	1.28%	0.52%	1.15%
R-Squared	0.9998	0.9989	0.9998	0.9992	0.9997	0.9991	0.9996	0.9992

The evaluation metrics derived from the XGBoost model's training dataset exhibit markedly lower values compared to those of the test dataset in both pre and post-feature selection scenarios, as illustrated in Table 5.7 and Table 5.8. This discrepancy implies a potential overfitting phenomenon within the training data. However, after the feature selection the error values for both the training and test datasets exhibit a reduction. This shift suggests a mitigated overfitting tendency within the training set, indicating a positive impact of feature selection on model generalization.

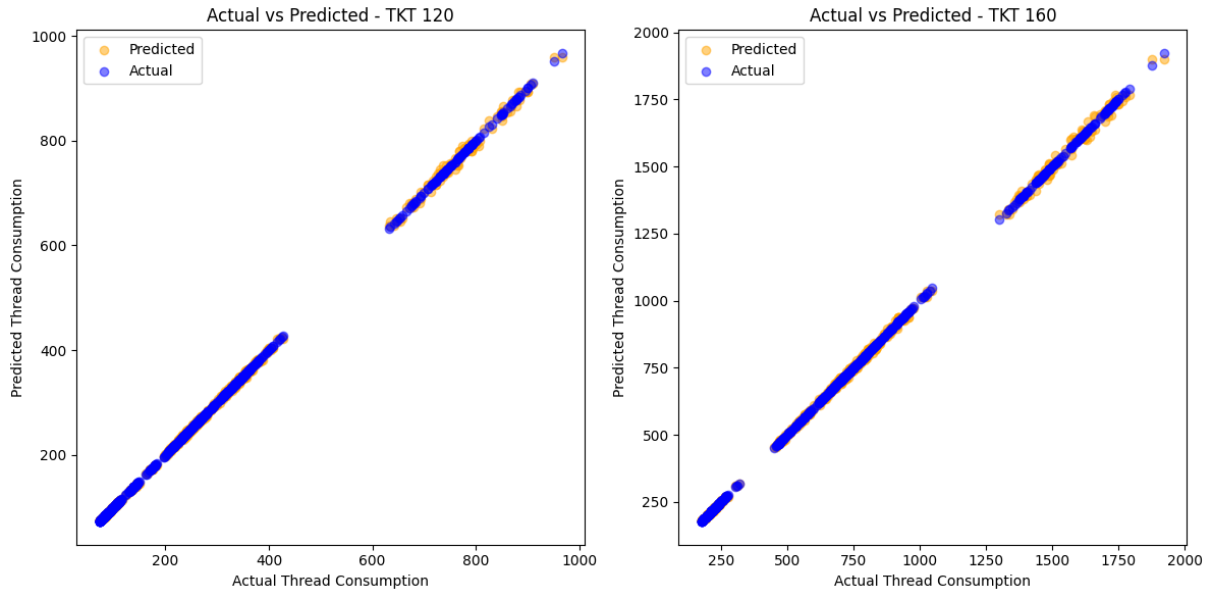


Figure 5.7 Actual vs Predicted values of XGBoost model

Slight deviations were seen in the actual vs predicted values in the XGBoost model for both thread types as demonstrated in the Figure 5.7.

Comparison of the Ensemble Models

The evaluation matrices retrieved from each ensemble model, post feature selection, have been consolidated in Table 5.9 below. A comparative analysis of each model's performance was conducted, revealing that XGBoost exhibited the lowest error values, indicating superior performance on the training dataset. Despite the notable proficiency demonstrated by XGBoost, a shared observation across all ensemble models studied suggests potential overfitting. This is evidenced by the discrepancy between low error values on the training dataset and comparatively higher error values on the test dataset, emphasizing the need for careful consideration of model generalization in future applications.

Table 5.9 Consolidated Evaluation matrices of Ensemble model – Consumption Model

Metrics	***Random Forest***				***Gradient Boosting***				***XGBoost***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	3.08	7.91	7.27	11.64	5.66	7.29	12.9	13.91	2.21	6.5	5.46	10.12
MAE	2.01	4.51	4.8	8.13	4.00	4.68	9.16	10.03	1.24	3.91	3.26	6.89
MAPE%	0.70	1.43	0.68	1.24	1.68	1.65	1.39	1.71	0.54	1.28	0.52	1.15

5.3 Multilayer Perceptron (Neural Network)

Neural networks with multiple output nodes can be used to predict the consumptions of multiple thread types. By configuring the output layer to have two nodes (one for TKT 120 and one for TKT 160), the model can simultaneously predict the wastages and consumptions of both thread types. Neural networks are known for their ability to capture complex patterns in data.

Initially the model consisted of only 3 layers namely, input layer, one hidden layer and one output layer with two outputs. The model was developed so that artificial neural networks were trained for TKT 120, with all the operations while for TKT 160, the operation which had zero consumption from TKT 160 was excluded from training. Moreover an additional hidden layer was added and the performance prior to and subsequent was recorded as shown in the Tables 5.10 and 5.11 for wastage prediction and consumption predictions respectively.

Table 5.10 Evaluation matrices of the ANN Models - Wastage Model

Metric for Wastage Model	***With 1 hidden layer***				***With 2 hidden layers***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	3.72	5.23	9.50	8.52	2.23	4.61	5.77	7.96
MAE	2.44	3.28	6.51	6.21	2.33	2.98	6.27	5.86
MAPE	52.43%	66.98%	56.86%	62.40%	53.89%	58.36%	61.14%	58.17%
R-Squared	0.9995	0.9986	0.9975	0.9968	0.9995	0.9996	0.9989	0.9998

Table 5.11 Evaluation matrices of the ANN Models - Consumption Model

Metrics for Consumption Model	***With 1 hidden layer***				***With 2 hidden layers***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	***Test***	***Train***	***Test***	***Train***	***Test***	***Train***	***Test***
RMSE	5.23	8.22	24.30	24.51	1.98	3.04	4.19	5.69
MAE	3.61	5.31	16.03	16.64	1.16	3.22	3.08	4.75
MAPE	1.45%	2.03%	2.60%	3.01%	1.15%	1.33%	1.99%	2.06%
R-Squared	0.9995	0.9986	0.9975	0.9968	0.9995	0.9996	0.9989	0.9998

As evidenced by the Table 5.10 and Table 5.11, incorporating an additional hidden layer has decreased the error values drastically hence enhancing the performance of the artificial neural network with 2 hidden layers additionally to the input and the output layers. Additionally, the low RMSE and MAPE values for training dataset suggest that models have trained well. MAPE values generated for the prediction of wastages lie around 52%-66% range which is comparatively a better indicator over the other models discussed.

The Artificial Neural Network model with 2 hidden layers was subjected to feature selection and excluded Operator Skill from the training. Evaluation matrices pre and post feature selection was included in the below Tables 5.12 and 5.13.

Table 5.12 Evaluation matrices of the ANN Wastage model pre and post feature selection

Metrics for Wastage Model	***With 2 hidden layers***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	***Test***	***Train***	***Test***	***Train***	***Test***	***Train***	***Test***
RMSE	2.23	4.61	5.77	7.96	1.95	2.98	3.89	4.79
MAE	2.33	2.98	6.27	5.86	1.23	2.64	3.44	4.56
MAPE	53.89%	58.36%	61.14%	58.17%	52.39%	57.71%	59.65%	59.95%
R-Squared	0.9995	0.9996	0.9989	0.9998	0.9995	0.9996	0.9989	0.9998

Table 5.13 Evaluation matrices of the ANN Consumption model pre and post feature selection

Metrics for Consumption Model	***With 2 hidden layers***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	***Test***	***Train***	***Test***	***Train***	***Test***	***Train***	***Test***
RMSE	1.98	3.04	4.19	5.69	1.56	2.67	3.41	4.37
MAE	1.16	3.22	3.08	4.75	1.08	2.98	2.82	3.95
MAPE	1.15%	1.33%	1.99%	2.06%	1.12%	1.21%	1.76%	1.99%
R-Squared	0.9995	0.9996	0.9989	0.9998	0.9995	0.9998	0.9999	0.9997

Feature selection has narrowed the MAPE value range of the wastage predictions to 50%-60% range. It was observed that the error values have reduced further more after the feature selection (Table 5.12 and Table 5.13) providing insights for the better model to be used. The same results were shown in the Figure 5.8 where the predicted values are very closer the the actual values.

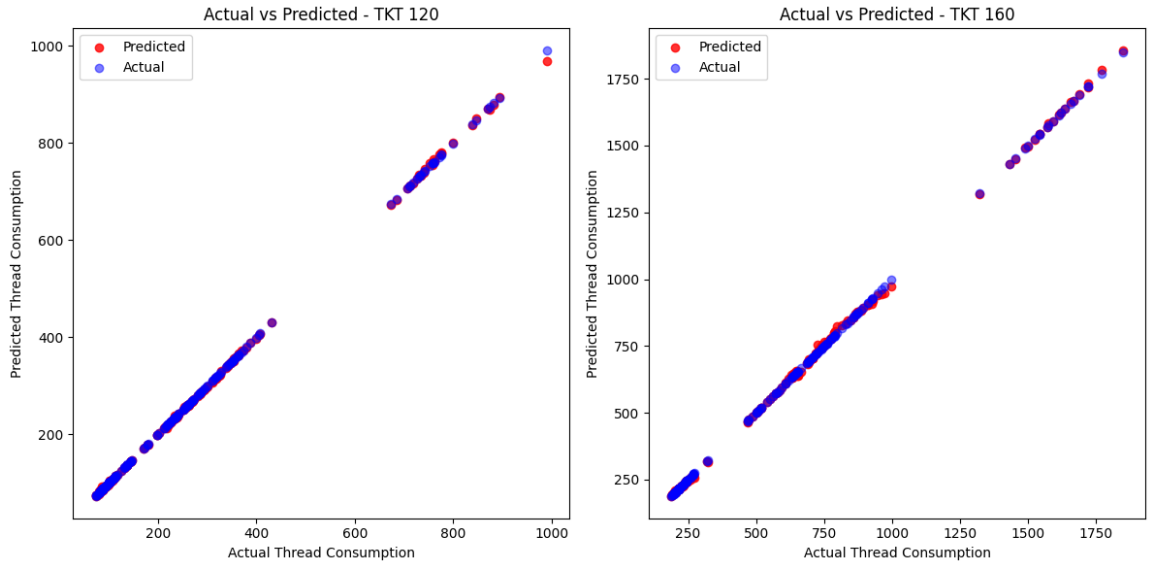


Figure 5.8 Actual vs Predicted values of ANN model

To avoid the overfitting phenomena, 'Early Stopping' technique was used in training artificial neural network models. Early stopping is a regularization technique used in machine learning and specifically in the training of iterative models, such as neural networks, to prevent overfitting and improve generalization. The fundamental idea behind early stopping is to track a model's performance on a validation set during training and halt the process when the model's performance starts to decline or stops getting better. The ANN model gave the best performance results among the other reviewed models as highlighted by the literature review. As per the studies ANN model was used to predict the thread consumption of only a single thread type to a specific stitch type where as in this study, ANN models were developed to predict the thread consumptions of two thread types in one whole operation as well as to predict the buffer wastage that should be incorporated to the operations inorder to balance any non-controllable wastages.

5.4 Development Of A User Interface

The development of a user interface for the study involved creating an interactive platform that seamlessly integrates machine learning models. In this context, Streamlit emerges as a powerful tool, offering a straightforward approach to building web applications using python. In the initial phase, the development environment was set up, ensuring that Python and Streamlit were properly installed. Subsequently, the necessary libraries were imported, including pandas and numpy for data manipulation, scikit-learn for machine learning functionalities, and Streamlit for building the interface.

The trained models which had best performances in predicting the thread consumptions (ANN models), specifically a model for TKT 120 and another for TKT 160, were loaded into the application. These models, implemented using libraries such as TensorFlow and Keras, were crucial for making predictions based on user inputs. This involved loading data from external sources, such as Excel files, and allowing users to input data through interactive forms in the Streamlit interface. Figure 5.9 is a prototype of the interface. The sewing operations were displayed using an expander for each operation to enhance the usability and the interactive nature of the application.

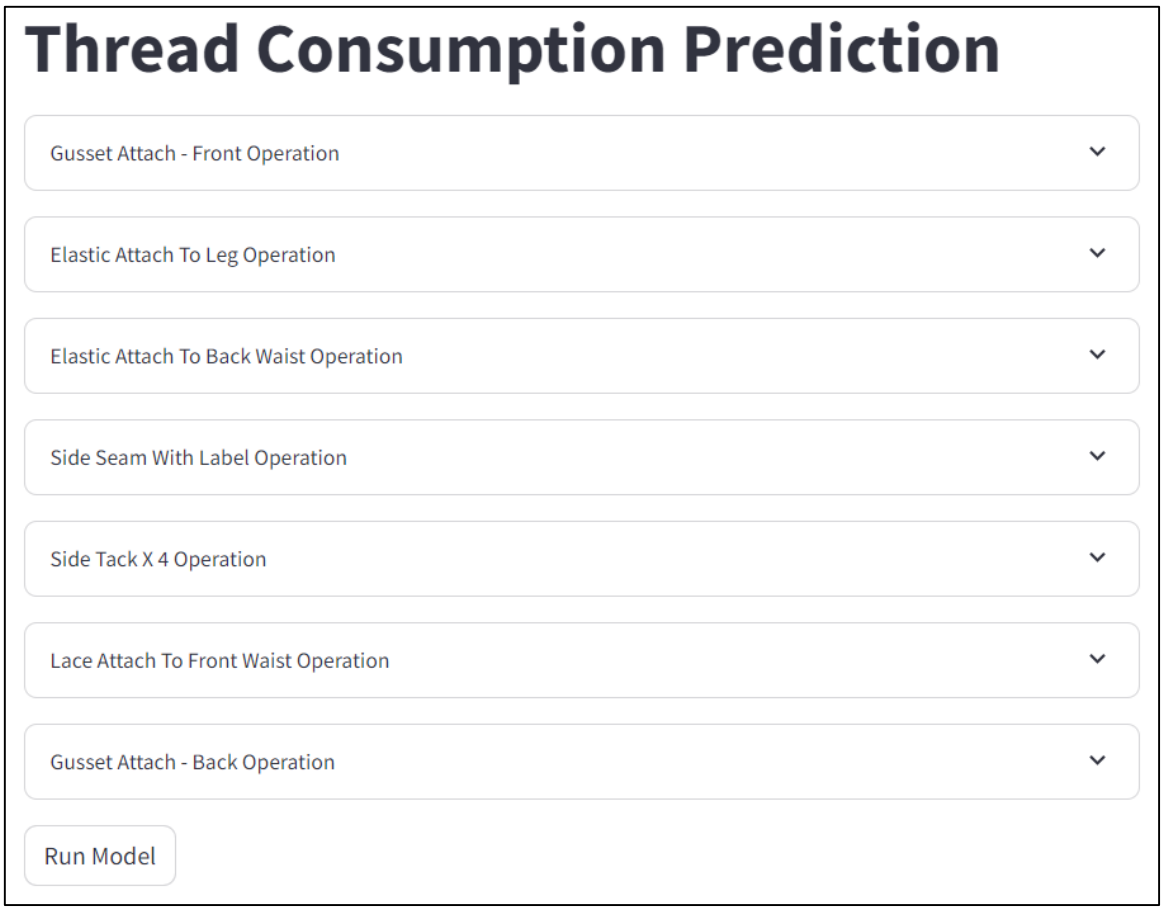


Figure 5.9 Proposed Interface for predicting thread consumption

The heart of the application lay in the creation of the Streamlit user interface. Utilizing Streamlit's features, various interactive elements like buttons, sliders, and select boxes were incorporated to enable users to input the parameters necessary for making predictions. Handling user inputs became a critical component. The application was designed to capture user inputs from the interface and process them into a format suitable for consumption by the machine learning models. This step involved defining how different input types, such as sliders and select boxes, would be interpreted and passed to the models. Figure 5.10 displays the view when an operation is expanded. It includes all the features such as Rows of stitch, SPI, Seam Length and Seam Thickness that were considered when the artificial neural network models were trained. The users need to carefully input the values for those features using this interface (Figure 5.10).

Figure 5.10 Expanded Operation - User Input Interface

Once the user inputs were processed, the application made predictions using the loaded machine learning models. For each operation and thread type (TKT 120 and TKT 160), the respective model was employed to predict thread wastage and the consumption based on the provided

parameters. The results of the model predictions were then displayed back to the users through the Streamlit interface. This included presenting the predicted values for thread consumption and any other relevant information, allowing users to gain insights into the outcomes based on their inputs.

During the development process, emphasis was placed on creating an intuitive and user-friendly interface. The layout was refined, and features were added based on feedback and usability testing. The goal was to ensure that users, regardless of their technical background, could easily interact with the application and comprehend the predictions generated by the machine learning models.

5.5 Chapter Summary

This chapter entailed training multiple models to forecast thread consumption in the apparel industry. The results indicate that the artificial neural network (ANN) model exhibited superior accuracy compared to other models and effectively addressed the problem of overfitting as evidenced by the literature.

After the best performing model was selected, an interactive user interface was developed for the prediction of thread consumption in a garment. The development of the user interface for this research using Streamlit provided a seamless platform for users to engage with and benefit from the machine learning models. The application succeeded in delivering an accessible and interactive experience, enabling users to make informed decisions based on the predictions generated by the TKT 120 and TKT 160 models.

CHAPTER 6
DISCUSSION, FUTURE WORK AND CONCLUSION

6.1 Discussion

The garment manufacturing industry is currently facing intensified competition, compelling organizations to prioritize cost control throughout the production process. A significant concern within this context is the accumulation of leftover stock, especially sewing thread, leading to increased write-off expenses. The increased inventory costs associated with storing and managing leftover thread stock pose financial challenges to the organization. The additional costs incurred for handling, storage, and transport directly affect profitability. Moreover, the organization's sustainability goals are at risk, as the disposal of leftover thread stock contributes to environmental concerns, including waste generation, pollution, and an elevated carbon footprint. Addressing these challenges is crucial for achieving efficient cost control and aligning with sustainability objectives.

The research narrows its focus to predict sewing thread consumption accurately, particularly in underwear fullbrief styles. By concentrating on a specific product category and stitch types, the study aims to comprehensively analyze the complexities associated with thread consumption prediction. The decision to focus on underwear fullbrief styles is strategic, considering their high production volume and relatively simpler construction. This allows for a detailed examination of factors influencing thread consumption within a specific and manageable scope. To enhance the accuracy in predicting thread consumption, the research employs statistical and machine learning techniques. Recognizing the multifactorial nature of thread utilization, the study considers variables such as garment style, fabric/seam thickness, stitch length, stitch density/SPI and seam type. The application of regression analysis and various machine learning models aimed to capture intricate patterns and trends in thread consumption, leading to more precise predictions.

Data required for the analysis were then collected and pre-processed to carry out a exploratory data analysis during which insightful findings were obtained. The hypothesis testing conducted focusing on whether the estimates for thread consumption, particularly for thread types TKT 120 and TKT 160 obtained from the product development team were overestimated, revealed that there is a statistically significant difference between estimated and actual thread consumptions for both thread types. The estimates were higher than the actual values which further emphasized the importance of an accurate thread consumption estimated method. The

buffer wastage inbuilt from the Product Development team appeared significantly higher than the actually measured wastages for TKT 120 and TKT 160, requiring a better approach in predicting the buffer wastages to be added to the consumption.

The explarotory data analysis of thread consumption data provided a solid foundation for the development of the predictive models considered in this study as the findings highlighted key operational parameters that significantly influence thread usage. As evidenced by the literature, the factors such SPI, Seam/fabric thickness, wastage and stictch type were identified as critical for the prediction of thread consumption from the findings of the explarotory data analysis. Additionally few more factors such as seam length, operator skill levels and operation type that appeared as influencial were considered as input parameters for the prediction models. Though seam length, wastages and stitch type were identified as factors affecting the thread consumption, none of the research work involved those factors in researches related to machine learning and neural network models to predict thread consumption. These have addressed the first research question under the Section 1.4 which inquired about the primary factors affecting the thread consumption in a garment and fulfilling the first objective under the section 1.3.2. Further in the analysis, the focus was shifted from predicting thread consumption based on specific stitch types to a more holistic approach. Instead of isolating predictions for individual stitch types, an operation wise context was focused. This departure from predicting thread usage for isolated stitches to predicting consumption for entire operations provided a more comprehensive consumption prediction for a specific product category.

To address the second research question under Section 1.4 of which statistical and machine learning models best suited for predicting thread consumption in a garment, the study analysed the performance of several statistical, machine learning and artificial neural network model predictions for the thread consumption. The insights to select the models were gained from the existing literature. The choice of the most suitable model depends on the characteristics of the dataset, the complexity of the relationships, and the specific requirements of the prediction task. It's essential to experiment with multiple models and evaluate their performance to determine which one provides the most accurate and reliable predictions for the thread consumptions of TKT 120 and TKT 160 in the given garment context. As in the literature review provided, different machine learning and neural network models will be trained to assure that the best machine learning model is adopted for predicting the actual thread consumption with the desired level of accuracy of 95% or above (Jaouadi, et al., 2006; Jaouachi & Khedher, 2013).

The exploration of various machine learning models, including Multivariate Linear Regression, Ensemble Models (Random Forest, Gradient Boosting, and XGBoost), and Multilayer Perceptron (Neural Network), revealed the influence of feature selection on predictive performance of both wastage prediction and consumption prediction. Through Multivariate Linear Regression, it became evident that separate models for each thread type and meticulous feature selection significantly enhanced predictive accuracy. Ensemble methods like Random Forest, Gradient Boosting, and XGBoost showcased significant responses to feature selection, with XGBoost emerging as the most adept yet susceptible to overfitting. The Multilayer Perceptron (Neural Network) model exhibited similar challenges but demonstrated remarkable improvement post-modification and feature selection, mitigating concerns of overfitting and enhancing overall predictive capability. These findings underscore the critical role of feature selection in refining model accuracy and generalization, addressing the third research question of how the choice of features influence the performance of machine learning models in predicting thread consumption.

In-lining with the first part of the forth reseach question of what metrics are most appropriate for evaluating the accuracy of machine learning models in predicting thread consumption, RMSE, MAE, MAPE and R-squared matrics enlightened by the existing literature, were used in the study to evaluate the performance of the models. The evaluation matrics were obtained for both the training and test data sets comprehensive assessment. While training set evaluation gauges the model's ability to learn from data, the test set evaluation assesses its capacity to generalize and make accurate predictions on new, unseen data. This practice helped to identify potential issues such as overfitting or underfitting, guided the selection of well-performing models, aided in hyperparameter tuning, and ensured the model's reliability across diverse scenarios, contributing to its overall effectiveness in real-world applications. In every model assessed, the R-squared metric consistently hovered around 99%, indicating its ineffectiveness for distinguishing between the models.

Multivariate Linear Regression

The initial attempt involved Multivariate Linear Regression, an extension of the simple linear regression model capable of handling multiple dependent variables. Feature selection was also carried out, but its impact on model performance was minimal as indicated by the insignificant variations in RMSE, MAE and MAPE values.

Ensemble Models

The exploration then extended to ensemble models, starting with Random Forest. While initially promising, there were signs of overfitting, emphasizing the need for careful consideration of generalization. Gradient Boosting and XGBoost were also employed, with XGBoost exhibiting the lowest error values but still showing signs of potential overfitting.

Multilayer Perceptron (Neral Network) Models

The application of Multilayer Perceptron (Neural Network) models for both wastage and total consumption prediction with one hidden layer outperformed the earlier models, multivariate linear regression and ensemble models. Moreover introducing an additional hidden layer and feature selection significantly improved the model's performance, reducing error values and providing valuable insights. The use of the 'Early Stopping' technique was also implemented to mitigate overfitting and showcase its effectiveness in enhancing model generalization.

The comparative analysis which was done to compare various prediction methods that have been emphasized in the literature, against the predictability of the proposed ANN model is shown in the Table 6.1. The Mean Absolute Percent Error (MAPE) is used to assess the prediction accuracy. As per the Table 6.1, geometrical models and regression models generally have higher MAPE values whereas artificial neural network models have lower MAPE values. This indicates that ANN models perform better in predicting thread consumptions. Moreover the lower MAPE values of the proposed ANN model exhibit enhaced performance out of the ANN models developed thus far.

Table 6.1 Comparison of different techniques of thread consumption prediction

	Geometrical Models			***Regression Models***			***Artificial Neural Network Models***					
Matric	***M1***	***M2***	***M3***	***M4***	***M5***	***M6***	***M7***	***M8***	***Proposed Model***			
									TKT 120		***TKT 160***	
									Train	***Test***	***Train***	***Test***
MAPE %	13.1	13.5	19.7	4.27	27.4	7.0	1.96	2.1	1.12	1.21	1.76	1.99
Notes: M1 - Model by (Ghosh & Chavhan, 2014), M2 - Model by (Jaouadi, et al., 2006), M3 - Model by (Chavan, et al., 2019), M4 - Model by (Jaouadi, et al., 2006), M5 - Model by (Abeysooriya & Wickramasinghe, 2014), M6 - Model by (Chavhan, et al., 2021), M7 – Model by (Jaouadi, et al., 2006) and M8 - Model by (Chavhan, et al., 2021)												

The study demonstrated the effectiveness of machine learning models in predicting thread consumptions, with the artificial neural network models showing superior accuracy as evidenced by the existing literature as well. This answers the second half of the fourth research question, which asks whether some machine learning models outperform others at predicting thread consumption. However, ongoing efforts are needed to address overfitting issues and enhance the generalization of these models in real-world garment industry applications. Future work should focus on refining model architectures, exploring additional features, and incorporating advanced regularization techniques to achieve more robust and reliable predictions.

The development of the user interface process focused on refining the interface layout, incorporating feedback, and ensuring a seamless experience for users. The user interface became a crucial tool for users to make informed decisions based on the predictions generated by the machine learning models. The application provided a platform for accurate thread consumption predictions, contributing to reduced write-off expenses, minimized inventory costs, and improved environmental sustainability.

6.2 Limitations

There are some limitations that should be taken into account when conducting this study. They are as follows.

I. **Time constraint:** This study is severely constrained by the 8 months time frame which restricts the amount of data that can be collected and examined. Due to the limited time, it may not be possible to gather a sufficient dataset to account for all possible variations in thread usage, which could have an influence on the accuracy and reliability of the model.

II. **Data limitations:** The accuracy and reliability of the prediction model will be affected by the volume and the quality of the data collected. The thread consumption worksheet is done considering only one size among different sizes ranging from S, M, L, XL, and 2XL. One base size is selected taking the size-wise quantities into account and the parameters for the relevant size of the particular style are measured, to avoid higher time consumption and workload involved in providing the thread consumption style-wise and size-wise. However, this would result in an overestimation of thread consumption for the sizes that are smaller than the base size and an underestimation for the sizes that are larger than the base size. To overcome this issue, a buffer wastage is added

considering the order quantities and size ratios and balances the thread consumption between sizes of the same style. This may result in an impact on the accuracy and reliability of the model.

III. **Model limitations:** The use of the proper features and algorithms will determine how well the prediction models function. Certain factors that affect sewing thread consumption could be difficult to measure or identify which could reduce the model's predictive power. For example tension of a thread is a significant factor that affects thread consumption but due to the unavailability of tension gauges in the factories, tension cannot be measured. Thus tension parameter is excluded from this study.

6.3 Suggestions For Future Work

First and foremost, addressing the limitations associated with data quality and volume is paramount. Enhancing the dataset by incorporating a broader array of sizes, styles, and product types can significantly bolster the accuracy and reliability of the prediction model. By considering size-wise and style-wise variations, the predictive model can offer more granular insights into the intricacies of different product configurations.

Additionally, the method for calculating wastage percentages needs refinement to better account for size variations. Developing a sophisticated algorithm that incorporates order quantities, size ratios, and other pertinent factors can contribute to a more accurate representation of thread consumption across varying sizes.

Further, it can be expanded beyond predictive models and delve into the development of thread optimization algorithms or models. While the current study has successfully tackled the challenge of predicting thread consumption, the next frontier lies in crafting algorithms that not only anticipate but also optimize the utilization of thread across different manufacturing products.

6.4 Conclusion

The main objective of this research was focused on predicting sewing thread consumption for two thread types, TKT 120 and TKT 160, across various sewing operations. In order to achieve this, the study meticulously addressed data preparation and preprocessing steps such as data cleaning, handling missing values, and data transformation and feature engineering.

The subsequent exploratory data analysis delved into the distribution and characteristics of the data. Hypothesis testing revealed that estimated thread consumption values were significantly greater than actual values, suggesting potential overestimation. Visualizations and statistical analyses, including scatter plots and regression, provided insights into the relationships between thread consumption, wastage percentages, stitches per inch (SPI), seam length, and seam thickness, unraveled insights to achieve the first sub objective of analyzing the key factors affecting thread consumption in a garment.

Second and third objectives of the study were successfully accomplished by employing machine learning techniques for prediction and evaluating their performance. The Multivariate Linear Regression model initially showed challenges, leading to model modifications and feature selection. Ensemble models such as Random Forest, Gradient Boosting, and XGBoost were introduced, each demonstrating varying degrees of accuracy and potential overfitting. Hence Multilayer Perceptron (Neural Network) appeared to be the best model for predicting thread consumptions as stated in the literature.

In summary, the research provides a comprehensive understanding of the complexities involved in predicting sewing thread consumption, incorporating meticulous data preparation, exploratory data analysis, and advanced analytics with machine learning techniques. The findings contribute valuable insights for industry practitioners aiming to optimize thread usage in garment manufacturing processes.

LIST OF REFERENCES

Ukponmwan, J. O., Mukhopadhyay, A. & Chatterjee, K. N., 2000. Sewing Threads. Textile Progress, 30(3), p. 91.

Abeysooriya, R. P. & Wickramasinghe, G. L. D., 2014. Regression model to predict thread consumption incorporating thread-tension constraint: study on lock-stitch 301 and chain-stitch 401.. Fashion and Textiles , Volume 1, pp. 1-8.

Abher, R. et al., 2014. Geometrical model to calculate the consumption of sewing thread for 301 Lockstitch.. The Journal of the Textile , Volume 105, pp. 1259-1264.

Abhishek, K., 2022. Introduction to artificial intelligence. [Online] Available at: https://www.red-gate.com/simple-talk/development/data-science-development/introduction-to-artificial-intelligence/
[Accessed 25 12 2023].

Amaan Group, 2010. Determining your sewing thread requirements. s.l.:Amann Group.

American & Efrid Inc, 2007. Technical bulletin estimating thread consumption.. s.l.:American & Efird Inc.

Barraza, N., Moro, S., Ferreyra, M. & De La Peña, A., 2018. Mutual information and sensitivity analysis for feature selection in customer targeting: A comparative study. Journal of Information Science, 45(1), pp. 53-67.

Breiman, L., 2001. Random Forests. Machine Learning, 45(1), p. 5–32.

Chauhan, R. & Ghosh, S., 2021. Geometric model of lockstitch seam and prediction of thread consumption. The Journal of The Textile Institute, Volume 113:2, pp. 314-323.

Chavan, M. V., Ghosh, S. & Naidu, M. R., 2019. An elliptical model for lockstitch 301 seam to estimate thread consumption.. The Journal of the Textile Institute, Volume 110(12), pp. 1740-1746.

Chavhan, M. V., Naidu, M. R. & Jamakhandi , H., 2021. Artificial neural network and regression models for prediction of sewing thread consumption for multilayered fabric assembly at lockstitch 301 seam. Research Journal of Textile and Apparel, 26(4), pp. 343-358.

Coats Digital, 2023. Coats SeamWorks. [Online] Available at: https://www.coats.com/en/solutions/coats-seamworks [Accessed 10 08 2023].

Copeland, B. J., 2023. Artificial intelligence | Definition, Examples, Types, Applications, Companies, & Facts.. [Online] Available at: https://www.britannica.com/technology/artificial-intelligence [Accessed 15 December 2023].

DOĞAN, S. & PAMUK, O., 2014. Calculating the amount of sewing thread consumption for different types of fabrics and stitch types. TEKSTİL ve KONFEKSİYON, 24(3).

Emjay International (Pvt) Ltd., 2023. FY 22/23 Book w-off, s.l.: s.n.

Emjay International and Penguin Sportswear, 2020. Emjay Penguin. [Online] Available at: http://www.emjayi.com/ [Accessed 02 April 2023].

Ghosh, S. & Chavhan, M. V., 2014. A geometrical model of stitch length for lockstitch seam. Indian Journal of Fibre and Textile Research,, 39(2), pp. 153-156.

Jaouachi , B. & Khedher , F., 2015. Evaluation of Sewed Thread Consumption of Jean Trousers Using Neural Network and Regression Methods. FIBRES & TEXTILES in Eastern Europe, 3(111), pp. 91-96.

Jaouachi , B. & Khedher, F., 2013. Evaluating sewing thread consumption of jean pants using fuzzy and regression methods.. The Journal of the Textile Institute, Volume 104, pp. 1065-1070.

Jaouachi, B. & Khedher, F., 2022. Assessment of jeans sewing thread consumption by applying metaheuristic optimization methods. International Journal of Clothing Science and Technology, 34(3), pp. 347-366.

Jaouadi, M., Msahli , S., Babay , A. & Zitouni , B., 2006. Analysis of the modelling methodologies for predicting the sewing thread consumption. International Journal of Clothing Science and Technology, Volume 18, pp. 7-18.

Khedher, F. & Jaouachi, B., 2015. Waste factor evaluation using theoretical and experimental jean pants consumptions. The Journal of The Textile Institute, 106(4), p. 402–408.

Mariem, B. et al., 2020. A Study of the Consumption of Sewing Threads for Women's Underwear: Bras and Panties. AUTEX Research Journal, 20(3), pp. 299-311.

Natekin, A. & Knoll, A., 2013. Gradient boosting machines, a tutorial. Frontiers in Neurorobotics, Volume 7.

Noriega, L., 2005. Multilayer Perceptron Tutorial.

Olive, D., 2017. Linear regression. s.l.:Springer.

Rasheed, A. et al., 2014. Geometrical model to calculate the consumption of sewing thread for 301 Lockstitch.. The Journal of the Textile Institute, 105(12), pp. 1259-1264.

Rehman, A. et al., 2021. Geometrical Model to Determine Sewing Thread Consumption for Stitch Class 406. FIBRES & TEXTILES in Eastern Europe , 29(6(150)), pp. 72-75.

Rengasamy, R. S. & Samuel, W. D., 2011. Effect of thread structure on tension peaks during lock stitch sewing. AUTEX Research Journal, 11(1), pp. 1-5.

Sarah, M., Boubaker, J., Faouzi, K. & Adolphe, D., 2020. Sewing thread consumption for different lockstitches of class 300 using geometrical and multi-linear regression models.. AUTEX Research Journal,, Volume 20, pp. 415-425.

Sharma , S., Gupta, V. & Midha, V. K., 2017. Predicting Sewing Thread Consumption for Chainstitch Using Regression. Journal of Textile Science & Engineering, 7(2), p. 295.

Vasiliev, V. A., Velmakina, Y. V. & Mayborodin, A. B., 2019. Using artificial neural networks when integrating the requirements of standards for management systems in QMS. IOP Conference Series: Materials Science and Engineering, 666(1), p. 012058.

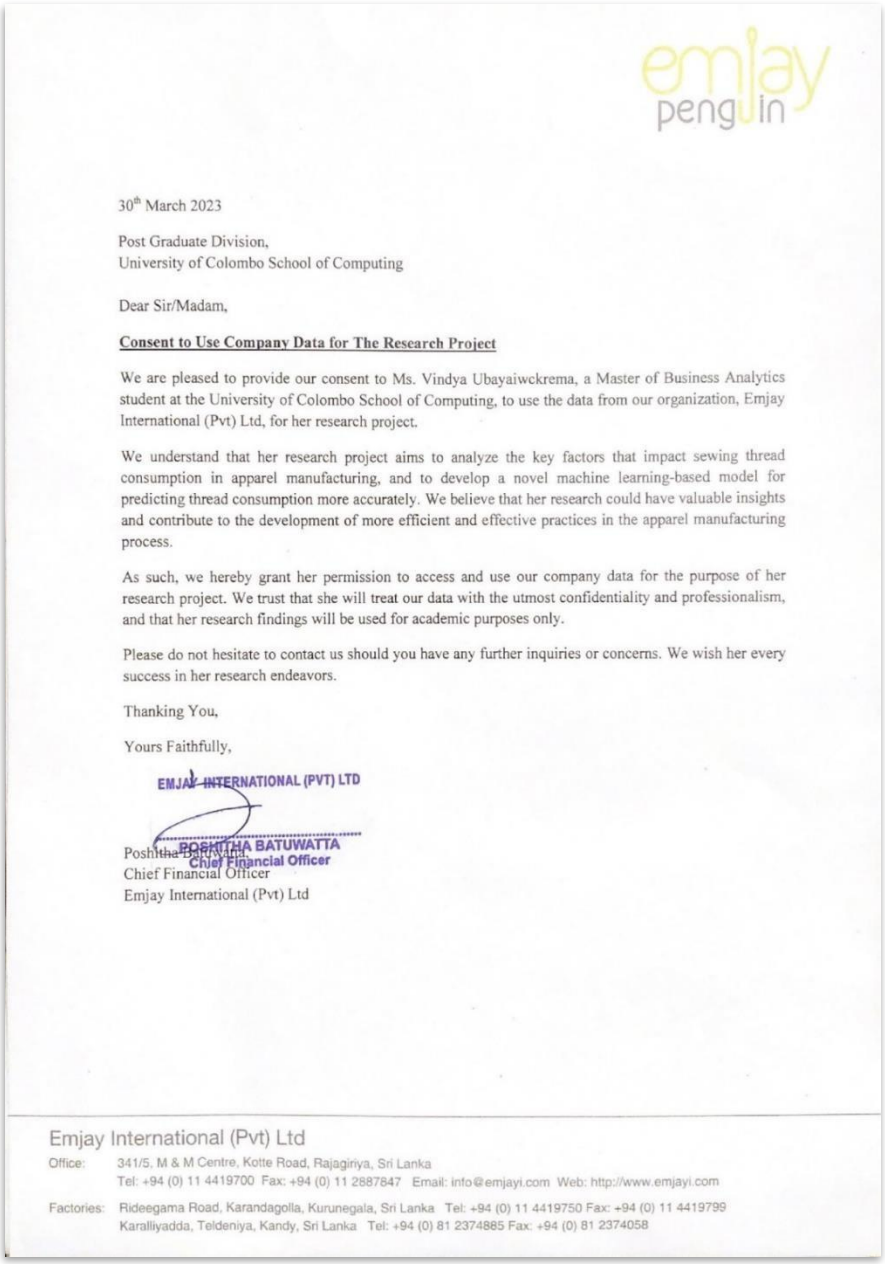
Wikipedia, 2023. Mutual information. [Online] Available at: https://en.wikipedia.org/wiki/Mutual_information [Accessed 8 December 2023].

Yeşilpınar, S. & Alkiraz, F., 2005. Kumaş kalınlığının dikiş iplik giderine etkisinin incelenmesi. The Journal of Textiles and Engineer, Volume 12(59-60), pp. 29-34..

APPENDICES

Appendix A : Supported Documents

A.1 Consent Letter



emjay penguin

30th March 2023

Post Graduate Division,
University of Colombo School of Computing

Dear Sir/Madam,

Consent to Use Company Data for The Research Project

We are pleased to provide our consent to Ms. Vindya Ubayaiwckrema, a Master of Business Analytics student at the University of Colombo School of Computing, to use the data from our organization, Emjay International (Pvt) Ltd, for her research project.

We understand that her research project aims to analyze the key factors that impact sewing thread consumption in apparel manufacturing, and to develop a novel machine learning-based model for predicting thread consumption more accurately. We believe that her research could have valuable insights and contribute to the development of more efficient and effective practices in the apparel manufacturing process.

As such, we hereby grant her permission to access and use our company data for the purpose of her research project. We trust that she will treat our data with the utmost confidentiality and professionalism, and that her research findings will be used for academic purposes only.

Please do not hesitate to contact us should you have any further inquiries or concerns. We wish her every success in her research endeavors.

Thanking You,

Yours Faithfully,

EMJAY INTERNATIONAL (PVT) LTD

POSHITHA BATUWATTA
Chief Financial Officer

Poshitha Batuwatta
Chief Financial Officer
Emjay International (Pvt) Ltd

Emjay International (Pvt) Ltd
Office: 341/5, M & M Centre, Kotte Road, Rajagiriya, Sri Lanka
Tel: +94 (0) 11 4419700 Fax: +94 (0) 11 2887847 Email: info@emjayi.com Web: http://www.emjayi.com
Factories: Rideegama Road, Karandagolla, Kurunegala, Sri Lanka Tel: +94 (0) 11 4419750 Fax: +94 (0) 11 4419799
Karalliyadda, Teldeniya, Kandy, Sri Lanka Tel: +94 (0) 81 2374885 Fax: +94 (0) 81 2374058

Appendix B: Samples of the Python Code

B.1 Sample code used to import data

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer


# Specify the path to your Excel file
excel_file_path = '/content/Thread.xlsx'

# Read the Excel file into a pandas DataFrame
df = pd.read_excel(excel_file_path)

# Now, 'df' contains your data from the Excel file
```

B.2 Sample code used to execute pre-processing steps

```
# Preprocessing and Feature Engineering steps
# Excluding "Cut Lace" Operation
df = df[df['Name of Operation'] != 'Cut Lace']

# Zero-Imputation for "B" and "L" Columns
df['Bobbin'].fillna(0, inplace=True)
df['Looper'].fillna(0, inplace=True)


# Transforming Symbols to Numerical Values
columns_to_transform = ['Needle Thread cm', 'Bobbin Thread cm', 'Looper Thread cm', 'TKT 160 COL 1']

for col in columns_to_transform:
    # Check if the column contains strings before using .str.replace
    if df[col].dtype == 'O':
        df[col] = df[col].str.replace(r'^\s*-\s*$', '0', regex=True)

# Omitting "Number of M/C Aallo." and "Stitch ID" Columns
df.drop(columns=['Number of M/C Aallo.', 'Stitch ID', 'Needle Thread', 'TKT 160 COL 1 Wastage cm','TKT 120 COL 1 Wastage cm',
                 'Bobbin Thread Tex / Type / Color', 'Looper Thread cm', 'Looper Thread Tex / Type / Color' ], inplace=True)

df['Total TKT 120 COL 1'] = df['TKT 120 COL 1'] + df['TKT 120 COL 1 Wastage cm']
df['Total TKT 160 COL 1'] = df['TKT 160 COL 1'] + df['TKT 160 COL 1 Wastage cm']
```

B.3 Sample code used in Multivariate Linear Regression modelling

```
# Extract features and targets for wastages
wastage_features = df_cleaned[['Operation','Rows of Stitch', 'SPI', 'Seam Length cm', 'Seam Thickness mm']]
wastage_targets = df_cleaned[['TKT 120 COL 1 Wastage cm', 'TKT 160 COL 1 Wastage cm']]

# Create a ColumnTransformer for wastages
wastage_preprocessor = ColumnTransformer(
    transformers=[
        ('num', 'passthrough', ['Rows of Stitch', 'SPI', 'Seam Length cm', 'Seam Thickness mm']),
        ('cat', OneHotEncoder(), ['Operation'])
    ])

# Fit and transform your data for wastages
X_wastage_transformed = wastage_preprocessor.fit_transform(wastage_features)

# Filter rows with 0 consumption for TKT 160 COL 1
non_zero_indices_tkt160 = df_cleaned[df_cleaned['Total TKT 160 COL 1'] != 0].index
filtered_df_tkt160 = df_cleaned.loc[non_zero_indices_tkt160]

# Extract features for TKT 160 wastages
tkt160_wastage_features = filtered_df_tkt160[['Operation', 'Rows of Stitch', 'SPI', 'Seam Length cm',
                                              'Seam Thickness mm']]
tkt160_wastage_targets = filtered_df_tkt160[['TKT 160 COL 1 Wastage cm']]

# Transform TKT 160 wastage features using the preprocessor
X_tkt160_wastage_transformed = wastage_preprocessor.transform(tkt160_wastage_features)

# Split the data into training and testing sets for TKT 160 wastages
X_tkt160_wastage_train, X_tkt160_wastage_test, y_tkt160_wastage_train, y_tkt160_wastage_test = train_test_split(
    X_tkt160_wastage_transformed, tkt160_wastage_targets, test_size=0.2, random_state=42)
```

```
# Apply the OneHotEncoder and train the model for TKT 120 wastages
tkt120_wastage_model = LinearRegression()
tkt120_wastage_model.fit(X_tkt120_wastage_train, y_tkt120_wastage_train)

# Transform TKT 120 consumption features using the preprocessor
X_tkt120_consumption_transformed = wastage_preprocessor.transform(tkt120_consumption_features)

# Split the data into training and testing sets for TKT 120 consumptions
X_tkt120_consumption_train, X_tkt120_consumption_test, y_tkt120_consumption_train, y_tkt120_consumption_test = train_test_split(
    X_tkt120_consumption_transformed, tkt120_consumption_targets, test_size=0.2, random_state=42)

# Train the model for TKT 120 consumptions
tkt120_consumption_model = LinearRegression()
tkt120_consumption_model.fit(X_tkt120_consumption_train, y_tkt120_consumption_train)

# Extract features for TKT 160 consumptions
tkt160_consumption_features = filtered_df_tkt160[['Operation','Rows of Stitch', 'SPI', 'Seam Length cm',
                                                  'Seam Thickness mm', 'TKT 160 COL 1 Wastage cm']]
tkt160_consumption_targets = filtered_df_tkt160[['Total TKT 160 COL 1']]

# Transform TKT 160 consumption features using the preprocessor
X_tkt160_consumption_transformed = wastage_preprocessor.transform(tkt160_consumption_features)

# Split the data into training and testing sets for TKT 160 consumptions
X_tkt160_consumption_train, X_tkt160_consumption_test, y_tkt160_consumption_train, y_tkt160_consumption_test = train_test_split(
    X_tkt160_consumption_transformed, tkt160_consumption_targets, test_size=0.2, random_state=42)

# Train the model for TKT 160 consumptions
tkt160_consumption_model = LinearRegression()
tkt160_consumption_model.fit(X_tkt160_consumption_train, y_tkt160_consumption_train)
```

B.4 Sample code used in Random Forest modelling

```
# Train the model for TKT 120 consumptions using RandomForestRegressor
tkt120_consumption_model = RandomForestRegressor(random_state=42)
tkt120_consumption_model.fit(X_tkt120_consumption_train, y_tkt120_consumption_train)

# Extract features for TKT 160 consumptions
tkt160_consumption_features = filtered_df_tkt160[['Operation', 'Operator Skill', 'Rows of Stitch', 'SPI',
                                                  'Seam Length cm', 'Seam Thickness mm', 'TKT 160 COL 1 Wastage cm']]
tkt160_consumption_targets = filtered_df_tkt160[['Total TKT 160 COL 1']]

# Transform TKT 160 consumption features using the preprocessor
X_tkt160_consumption_transformed = wastage_preprocessor.transform(tkt160_consumption_features)

# Split the data into training and testing sets for TKT 160 consumptions
X_tkt160_consumption_train, X_tkt160_consumption_test, y_tkt160_consumption_train, y_tkt160_consumption_test = train_test_split(
    X_tkt160_consumption_transformed, tkt160_consumption_targets, test_size=0.2, random_state=42)

# Train the model for TKT 160 consumptions using RandomForestRegressor
tkt160_consumption_model = RandomForestRegressor(random_state=42)
tkt160_consumption_model.fit(X_tkt160_consumption_train, y_tkt160_consumption_train)
```

```
<ipython-input-16-d43dd966edd8>:63: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please chang
  tkt120_wastage_model.fit(X_tkt120_wastage_train, y_tkt120_wastage_train)
<ipython-input-16-d43dd966edd8>:104: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please char
  tkt120_consumption_model.fit(X_tkt120_consumption_train, y_tkt120_consumption_train)
<ipython-input-16-d43dd966edd8>:118: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please char
  tkt160_consumption_model.fit(X_tkt160_consumption_train, y_tkt160_consumption_train)
```

```
RandomForestRegressor
RandomForestRegressor(random_state=42)
```

B.5 Sample code used in Gradient Boosting modelling

```
# Train the model for TKT 120 consumptions using GradientBoostingRegressor
tkt120_consumption_model = GradientBoostingRegressor(random_state=42)
tkt120_consumption_model.fit(X_tkt120_consumption_train, y_tkt120_consumption_train)

# Extract features for TKT 160 consumptions
tkt160_consumption_features = filtered_df_tkt160[['Operation', 'Operator Skill', 'Rows of Stitch', 'SPI', 'Seam Length cm',
                                                  'Seam Thickness mm', 'TKT 160 COL 1 Wastage cm']]
tkt160_consumption_targets = filtered_df_tkt160[['Total TKT 160 COL 1']]

# Transform TKT 160 consumption features using the preprocessor
X_tkt160_consumption_transformed = wastage_preprocessor.transform(tkt160_consumption_features)

# Split the data into training and testing sets for TKT 160 consumptions
X_tkt160_consumption_train, X_tkt160_consumption_test, y_tkt160_consumption_train, y_tkt160_consumption_test = train_test_split(
    X_tkt160_consumption_transformed, tkt160_consumption_targets, test_size=0.2, random_state=42)

# Train the model for TKT 160 consumptions using GradientBoostingRegressor
tkt160_consumption_model = GradientBoostingRegressor(random_state=42)
tkt160_consumption_model.fit(X_tkt160_consumption_train, y_tkt160_consumption_train)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_gb.py:437: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_gb.py:437: DataConversionWarning: A column-vector y was passed when a
  y = column_or_1d(y, warn=True)
```

```
GradientBoostingRegressor
GradientBoostingRegressor(random_state=42)
```

B.6 Sample code used in XGBoost modelling

```
# Split the data into training and testing sets for TKT 120 consumptions
X_tkt120_consumption_train, X_tkt120_consumption_test, y_tkt120_consumption_train, y_tkt120_consumption_test = train_test_split(
    X_tkt120_consumption_transformed, tkt120_consumption_targets, test_size=0.2, random_state=42)

# Train the model for TKT 120 consumptions using XGBRegressor
tkt120_consumption_model = XGBRegressor(random_state=42)
tkt120_consumption_model.fit(X_tkt120_consumption_train, y_tkt120_consumption_train)

# Extract features for TKT 160 consumptions
tkt160_consumption_features = filtered_df_tkt160[['Operation', 'Operator Skill', 'Rows of Stitch', 'SPI', 'Seam Length cm',
                                                  'Seam Thickness mm', 'TKT 160 COL 1 Wastage cm']]
tkt160_consumption_targets = filtered_df_tkt160[['Total TKT 160 COL 1']]

# Transform TKT 160 consumption features using the preprocessor
X_tkt160_consumption_transformed = wastage_preprocessor.transform(tkt160_consumption_features)

# Split the data into training and testing sets for TKT 160 consumptions
X_tkt160_consumption_train, X_tkt160_consumption_test, y_tkt160_consumption_train, y_tkt160_consumption_test = train_test_split(
    X_tkt160_consumption_transformed, tkt160_consumption_targets, test_size=0.2, random_state=42)

# Train the model for TKT 160 consumptions using XGBRegressor
tkt160_consumption_model = XGBRegressor(random_state=42)
tkt160_consumption_model.fit(X_tkt160_consumption_train, y_tkt160_consumption_train)
```

```
XGBRegressor
XGBRegressor(base_score=None, booster=None, callbacks=None,
             colsample_bylevel=None, colsample_bynode=None,
             colsample_bytree=None, device=None, early_stopping_rounds=None,
             enable_categorical=False, eval_metric=None, feature_types=None,
             gamma=None, grow_policy=None, importance_type=None,
```

B.7 Sample code used in ANN modelling

```
# Build the ANN model for TKT 120
tkt120_consumption_model = Sequential()
tkt120_consumption_model.add(Dense(128, activation='relu', input_shape=(X_tkt120_consumption_train_combined.shape[1],)))
tkt120_consumption_model.add(Dense(64, activation='relu'))
tkt120_consumption_model.add(Dense(32, activation='relu'))
tkt120_consumption_model.add(Dense(1, activation='linear'))

# Compile the model for TKT 120
tkt120_consumption_model.compile(optimizer='adam', loss='mean_squared_error', metrics=['mae', 'mape'])

# Build the ANN model for TKT 160
tkt160_consumption_model = Sequential()
tkt160_consumption_model.add(Dense(128, activation='relu', input_shape=(X_tkt160_consumption_train_combined.shape[1],)))
tkt160_consumption_model.add(Dense(64, activation='relu'))
tkt160_consumption_model.add(Dense(32, activation='relu'))
tkt160_consumption_model.add(Dense(1, activation='linear'))

# Compile the model for TKT 160
tkt160_consumption_model.compile(optimizer='adam', loss='mean_squared_error', metrics=['mae', 'mape'])

# Train the consumption models
# Train the model for TKT 120
history_tkt_120 = tkt120_consumption_model.fit(
    X_tkt120_consumption_train_combined, y_train_tkt_120,
    epochs=500, batch_size=32,
    validation_split=0.2, verbose=2,
    callbacks=[EarlyStopping(patience=10, restore_best_weights=True)]
)
```

B.8 Sample code used for Evaluation Matrics of Wastage Model

```
from sklearn.metrics import mean_squared_error, mean_absolute_error

# Make predictions on the train sets
Wastage_predictions_tkt_120 = tkt120_wastage_model.predict(X_tkt120_wastage_train)
Wastage_predictions_tkt_160 = tkt160_wastage_model.predict(X_tkt160_wastage_train)

# Evaluate the model performance
mse_tkt_120_wastage = mean_squared_error(y_tkt120_wastage_train, Wastage_predictions_tkt_120)
mse_tkt_160_wastage = mean_squared_error(y_tkt160_wastage_train, Wastage_predictions_tkt_160)

print(f'Mean Squared Error for TKT 120 wastage: {mse_tkt_120_wastage}')
print(f'Mean Squared Error for TKT 160 wastage: {mse_tkt_160_wastage}')

# Calculate RMSE for both thread types
rmse_tkt_120_wastage = np.sqrt(mse_tkt_120_wastage)
rmse_tkt_160_wastage = np.sqrt(mse_tkt_160_wastage)

print(f'Root Mean Squared Error for TKT 120 wastage: {rmse_tkt_120_wastage}')
print(f'Root Mean Squared Error for TKT 160 wastage: {rmse_tkt_160_wastage}')


# Calculate MAE for both thread types
mae_tkt_120_wastage = mean_absolute_error(y_tkt120_wastage_train, Wastage_predictions_tkt_120)
mae_tkt_160_wastage = mean_absolute_error(y_tkt160_wastage_train, Wastage_predictions_tkt_160)

print(f'Mean Absolute Error for TKT 120 wastage: {mae_tkt_120_wastage}')
print(f'Mean Absolute Error for TKT 160 wastage: {mae_tkt_160_wastage}')
```

```
def calculate_mape(y_true, y_pred):
    """
    Calculate Mean Absolute Percentage Error (MAPE).

    Parameters:
    - y_true: array-like, shape (n_samples,) - Ground truth (correct) values.
    - y_pred: array-like, shape (n_samples,) - Predicted values.

    Returns:
    - mape: float - Mean Absolute Percentage Error.
    """
    assert len(y_true) == len(y_pred), "Input arrays must have the same length."

    # Calculate absolute percentage error
    absolute_percentage_error = np.abs((y_true - y_pred) / y_true)

    # Exclude entries where y_true is zero
    absolute_percentage_error = absolute_percentage_error[y_true != 0]

    # Calculate MAPE
    mape = np.mean(absolute_percentage_error) * 100

    return mape

# Calculate MAPE for both thread types
mape_tkt120_wastage = calculate_mape(y_tkt120_wastage_train.values.flatten(), Wastage_predictions_tkt_120.flatten())
mape_tkt160_wastage = calculate_mape(y_tkt160_wastage_train.values.flatten(), Wastage_predictions_tkt_160.flatten())

print(f'MAPE for TKT 120 wastage: {mape_tkt120_wastage:.2f}%')
print(f'MAPE for TKT 160 wastage: {mape_tkt160_wastage:.2f}%')
```

```
from sklearn.metrics import r2_score
r2_tkt_120_wastage = r2_score(y_tkt120_wastage_train, Wastage_predictions_tkt_120)
r2_tkt_160_wastage = r2_score(y_tkt160_wastage_train, Wastage_predictions_tkt_160)

print(f'R-squared for TKT 120 wastage: {r2_tkt_120_wastage}')
print(f'R-squared for TKT 160 wastage: {r2_tkt_160_wastage}')
```

B.9 Sample code used for Evaluation Matrics of Consumption Model

```
#Evaluating the model for consumptions
# Make predictions on the train sets
Consumption_predictions_tkt_120 = tkt120_consumption_model.predict(X_tkt120_consumption_train)
Consumption_predictions_tkt_160 = tkt160_consumption_model.predict(X_tkt160_consumption_train)

# Evaluate the model performance
mse_tkt_120_consumption = mean_squared_error(y_tkt120_consumption_train, Consumption_predictions_tkt_120)
mse_tkt_160_consumption = mean_squared_error(y_tkt160_consumption_train, Consumption_predictions_tkt_160)

print(f'Mean Squared Error for TKT 120 consumption: {mse_tkt_120_consumption}')
print(f'Mean Squared Error for TKT 160 consumption: {mse_tkt_160_consumption}')

# Calculate RMSE for both thread types
rmse_tkt_120_consumption = np.sqrt(mse_tkt_120_consumption)
rmse_tkt_160_consumption = np.sqrt(mse_tkt_160_consumption)

print(f'Root Mean Squared Error for TKT 120 consumption: {rmse_tkt_120_consumption}')
print(f'Root Mean Squared Error for TKT 160 consumption: {rmse_tkt_160_consumption}')


# Calculate MAE for both thread types
mae_tkt_120_consumption = mean_absolute_error(y_tkt120_consumption_train, Consumption_predictions_tkt_120)
mae_tkt_160_consumption = mean_absolute_error(y_tkt160_consumption_train, Consumption_predictions_tkt_160)

print(f'Mean Absolute Error for TKT 120 consumption: {mae_tkt_120_consumption}')
print(f'Mean Absolute Error for TKT 160 consumption: {mae_tkt_160_consumption}')
```

```
def calculate_mape(y_true, y_pred):
    """
    Calculate Mean Absolute Percentage Error (MAPE).

    Parameters:
    - y_true: array-like, shape (n_samples,) - Ground truth (correct) values.
    - y_pred: array-like, shape (n_samples,) - Predicted values.

    Returns:
    - mape: float - Mean Absolute Percentage Error.
    """
    assert len(y_true) == len(y_pred), "Input arrays must have the same length."

    # Calculate absolute percentage error
    absolute_percentage_error = np.abs((y_true - y_pred) / y_true)

    # Exclude entries where y_true is zero
    absolute_percentage_error = absolute_percentage_error[y_true != 0]

    # Calculate MAPE
    mape = np.mean(absolute_percentage_error) * 100

    return mape

# Calculate MAPE for both thread types
mape_tkt120_consumption = calculate_mape(y_tkt120_consumption_train.values.flatten(), Consumption_predictions_tkt_120.flatten())
mape_tkt160_consumption = calculate_mape(y_tkt160_consumption_train.values.flatten(), Consumption_predictions_tkt_160.flatten())

print(f'MAPE for TKT 120 consumption: {mape_tkt120_consumption:.2f}%')
print(f'MAPE for TKT 160 consumption: {mape_tkt160_consumption:.2f}%')
```

```
from sklearn.metrics import r2_score
r2_tkt_120_consumption = r2_score(y_tkt120_consumption_train, Consumption_predictions_tkt_120)
r2_tkt_160_consumption = r2_score(y_tkt160_consumption_train, Consumption_predictions_tkt_160)

print(f'R-squared for TKT 120 consumption: {r2_tkt_120_consumption}')
print(f'R-squared for TKT 160 consumption: {r2_tkt_160_consumption}')
```

B.10 Sample code used to create the User Interface

```
# Streamlit app
def main():
    global user_inputs  # Use the global variable

    st.title("Thread Consumption Prediction")

    # Get unique operations
    unique_operations = df_cleaned['Operation'].unique()

    # Create a separate column for each operation
    for operation in unique_operations:
        with st.expander(f"{operation} Operation"):
            # Get the image path for the current operation from the dictionary
            image_path = operation_images.get(operation, '/content/Full Brief Style.jpg')  # Provide a default image path
            image = Image.open(image_path)

            # User input section for the specific operation
            col1, col2, col3, col4, col5 = st.columns(5)
            rows_of_stitch = col1.selectbox(f"Rows of Stitch", ["1", "2", "4"], key=f"{operation}_rows_of_stitch")
            spi = col2.text_input(f"SPI", key=f"{operation}_spi")
            seam_length_cm = col3.text_input(f"Seam Length cm", key=f"{operation}_seam_length")
            seam_thickness_mm = col4.text_input(f"Seam Thickness mm", key=f"{operation}_seam_thickness")
```

```
            user_inputs = user_inputs.append({
                'Operation': operation,
                'Rows of Stitch': rows_of_stitch,
                'SPI': spi,
                'Seam Length cm': seam_length_cm,
                'Seam Thickness mm': seam_thickness_mm,
            }, ignore_index=True)
            image_location = st.empty()
            image_location.image(image, use_column_width=False, width=500, clamp=True, output_format="auto")


    # Prediction button
    if st.button("Run Model"):
        # Transform user input for TKT 120
        X_user_input_tkt_120 = preprocessor_tkt_120_wastages.transform(user_inputs)
        X_user_input_tkt_120_scaled = scaler_tkt_120_wastages.transform(X_user_input_tkt_120)

        # Transform user input for TKT 160 (excluding 'Side Tack X 4')
        X_user_input_tkt_160 = preprocessor_tkt_160_wastages.transform(user_inputs[user_inputs['Operation'] != 'Side Tack X 4'])
        X_user_input_tkt_160_scaled = scaler_tkt_160_wastages.transform(X_user_input_tkt_160)

        # Predict thread wastages for TKT 120
        predictions_tkt_120_wastages = tkt120_wastage_model.predict(X_user_input_tkt_120_scaled)
        total_wastage_tkt_120 = np.sum(predictions_tkt_120_wastages)

        # Predict thread wastages for TKT 160
        predictions_tkt_160_wastages = tkt160_wastage_model.predict(X_user_input_tkt_160_scaled)
        total_wastage_tkt_160 = np.sum(predictions_tkt_160_wastages)
```

```
        # Reshape the predicted wastages to (n_samples, 1)
        predictions_tkt_120_wastages = predictions_tkt_120_wastages.reshape(-1, 1)
        predictions_tkt_160_wastages = predictions_tkt_160_wastages.reshape(-1, 1)

        # Concatenate predicted wastages to user input features
        X_user_input_tkt_120_combined = np.concatenate([X_user_input_tkt_120_scaled, predictions_tkt_120_wastages], axis=1)
        X_user_input_tkt_160_combined = np.concatenate([X_user_input_tkt_160_scaled, predictions_tkt_160_wastages], axis=1)

        # Predict thread consumption for TKT 120
        predictions_tkt_120 = tkt120_consumption_model.predict(X_user_input_tkt_120_combined)
        total_consumption_tkt_120 = np.sum(predictions_tkt_120)

        # Predict thread consumption for TKT 160
        predictions_tkt_160 = tkt160_consumption_model.predict(X_user_input_tkt_160_combined)
        total_consumption_tkt_160 = np.sum(predictions_tkt_160)

        st.write("#### TKT 120:")
        st.success(f"Predicted Total Thread Wastage for TKT 120: {total_wastage_tkt_120:.2f} cm")
        st.success(f"Predicted Total Thread Consumption for TKT 120: {total_consumption_tkt_120:.2f} cm")

        st.write("#### TKT 160:")
        st.success(f"Predicted Total Thread wastage for TKT 160: {total_wastage_tkt_160:.2f} cm")
        st.success(f"Predicted Total Thread Consumption for TKT 160: {total_consumption_tkt_160:.2f} cm")

if __name__ == "__main__":
    main()
```

Appendix C: Drafted Research Paper

Analyse Factors Affecting Thread Consumption in a Garment and Develop a Machine Learning-based Prediction Model

Abstract

The garment manufacturing industry faces intensified competition, prompting the need for cost control and efficient inventory management. This research addresses the challenges of excess thread stock, leading to increased write-off expenses and environmental concerns. Focusing on predicting sewing thread consumption in underwear fullbrief styles, the study employs statistical and machine learning techniques, considering variables such as garment style, fabric/seam thickness, stitch length, stitch density/SPI, seam type, and estimated wastage.

The development of a user interface using Streamlit integrates machine learning models for two types of threads, allowing users to input parameters through an intuitive layout. The user-friendly interface facilitates informed decision-making based on predictions of total thread consumption. The application contributes to reducing write-off expenses, minimizing inventory costs, and aligning with environmental sustainability goals.

The research highlights the effectiveness of machine learning models, particularly artificial neural network models, in predicting thread consumption. Overcoming challenges such as overfitting and enhancing generalization, the study emphasizes the need for refining model architectures and exploring additional features. The user interface development emerges as a crucial tool for achieving efficient cost control and sustainability in the garment manufacturing industry.

Keywords: Garment Manufacturing, Thread Consumption Prediction, Machine Learning, Artificial Neural Network, User Interface, Streamlit, Cost Control.

1. Introduction

The garment manufacturing industry faces intensified competition, prompting organizations to minimize costs from material procurement to order completion. Raw material expenses significantly impact the contribution margin, including actual material usage, waste, and leftover stock. The organization subjected to this study, has excess material, termed as write-off stock, which amounts to $1,479,457, with fabric and sewing thread being major components. In the financial year 2022/2023, the organization's write-off rate exceeded the company standard of 2%, hitting 4.1%, increasing costs, lowering profitability, and accumulating excess inventory.

This research aims to minimize leftover sewing thread stock, valued at $75,059, through accurate thread consumption prediction. The problem identified lies in excessive leftover sewing thread stock postproduction, indicating potential overestimation of thread requirements, leading to increased inventory costs and environmental impact. The research seeks to address these challenges by analysing factors influencing thread consumption and developing an accurate prediction model. Enhancing accuracy in thread consumption prediction is vital for optimal resource usage and reducing unused stock on the production floor. Thread consumption varies based on garment styles, sizes, fabrics, stitch parameters, and seam types. Traditional estimation methods lack precision, prompting exploration into machine learning models. By developing a machine learning-based prediction model, the aim is to precisely estimate thread quantities, reduce unused stock, and enhance profitability. Objectives include analysing key factors affecting thread consumption, applying various machine

learning models, evaluating model accuracy, and developing a user-friendly interface for thread consumption prediction.

2. Literature Review

The literature review explores research initiatives related to the study, focusing on factors influencing sewing thread consumption in garment manufacturing and the development of machine learning-based models.

Factors affecting thread consumption include garment type, size, design, fabric, stitch length, stitch density, and stitch type (Ukponmwan, et al., 2000). These variables impact the amount of thread used, with considerations such as heavier fabrics and longer stitch lengths increasing consumption. Yeşilpınar and Alkiraz (2005) investigated the impact of fabric thickness on sewing thread consumption, revealing a direct correlation between fabric thickness and thread usage. Thicker fabrics demand more thread for a durable seam, with fabric thickness identified as the most significant factor influencing thread consumption. The type of stitch used also affects consumption, with lockstitch consuming the most, followed by chain stitch, three-thread overedge stitch, and four-thread overedge stitch (Yeşilpınar & Alkiraz, 2005). Rengasamy and Samuel (2011) highlighted the importance of thread tension in garment construction, impacting the quantity of thread used for seams. Incorrect thread tension can result in puckered seams, thread breakage, or unravelling, leading to increased thread usage. Understanding and managing thread tension are crucial for optimizing thread consumption and producing high-quality seams (Rengasamy & Samuel, 2011).

Researchers have sought alternative prediction methods, including value prediction charts, mathematical formulas, thread length ratios, predictive algorithms, learning algorithms, and software solutions (Jaouadi, et al., 2006). Garment manufacturers traditionally used graphs, tables, and formulas based on assumptions and trial-and-error methods (American & Efrid Inc, 2007; Amaan Group, 2010). However, these methods lacked flexibility with varying fabric thickness and stitch densities, raising questions about predicted values' accuracy. Consumption ratios, initially limited to one stitch density value, were customized for different stitches by thread suppliers, allowing accurate calculation of thread usage considering various parameters (American & Efrid Inc, 2007; Amaan Group, 2010). Leading thread suppliers now use software packages to enhance accuracy, employing formulas and ratios for diverse parameters like stitch lengths, densities, and fabric thicknesses (Abeysooriya & Wickramasinghe, 2014). For instance, Coats introduced SEAMWORKS, a software calculating sewing thread amount based on parameters like stitch types and color groups, providing a detailed result report with the total cost for the used sewing thread (Coats Digital, 2023).

Researchers have addressed issues with traditional methods like graphs, tables, and formulas by proposing mathematical and geometrical models. Some studies have considered the geometric shape of stitch types, such as the 301 lockstitch, and developed models with high accuracy, emphasizing factors like stitch width, density, and needle distance (Rasheed, et al., 2014).

Geometric models based on rectangular profiles have been prevalent, but recent advancements include models with realistic elliptical profiles, demonstrating lower error rates and better generalization across fabric types and densities (Chauhan & Ghosh, 2021). Regression models incorporating factors like thread tension, fabric thickness, and stitch density have been proposed, offering improved accuracy and reduction in error percentages for specific stitch types (Abeysooriya & Wickramasinghe, 2014).

Machine learning, neural network models, and metaheuristic optimization models have gained prominence. Artificial neural network models, in particular, have shown high reliability in predicting thread consumption, achieving an accuracy of at least 95% (Jaouadi, et al., 2006). Fuzzy theory-

based models offer flexibility, considering parameters like thread composition, needle size, and fabric weight (Jaouachi & Khedher, 2013). Studies comparing analytical models, regression models, and artificial neural network models highlight the superior performance of the artificial neural network model in estimating thread consumption.

Additionally, metaheuristic optimization techniques, including particular swarm optimization (PSO), ant colony optimization (ACO), and genetic algorithm (GA), have been explored to minimize thread consumption (Jaouachi & Khedher, 2022). Results indicate that PSO and ACO methods are more precise than experimental methods, showcasing the potential for optimization in thread usage.

The literature review underscores the evolution from traditional methods to more sophisticated models, emphasizing the superior performance of machine learning-based approaches in accurately predicting sewing thread consumption. The study aims to contribute to this progress by analysing new variables affecting thread consumption and developing a novel initiative using machine learning techniques to enhance accuracy in predicting thread consumption in garment manufacturing.

3. Methodology and Theory

In particular, the research focuses on predicting thread consumption in the fullbrief styles of underwear in size Medium. Fullbrief styles were selected due to their high production volume and simpler construction with fewer operations. The study will evaluate three commonly used stitch types (301, 406, and 514) for a particular fullbrief style (Figure 3.1), using two different thread types (TKT 120 and TKT 160). TKT (Ticket) numbers is used to determine the thread's thickness or linear density. The TKT number is the number of thousands of yards (or length of thread) needed to weigh one pound. In general, thinner or finer threads are indicated by higher TKT values, whereas thicker threads are indicated by lower TKT numbers.

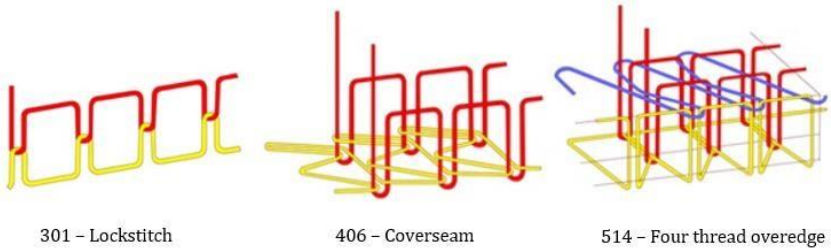


Figure 3.1 Different types of Stitches used in the study

In order to measure and track the amount of wastage for each type of thread used in each operation, a cooperative effort was formed with the Business Process (BP) team. Additionally, the study included a human component that acknowledged the critical role that machine operators have in determining how production processes turn out. Machine Operator (MO) Grading Report, a dedicated database, was used to help record operator skill levels in the thread consumption database. The collected data were then pre-processed and subjected to feature engineering for further analysis.

4. Exploratory Data Analysis

Exploratory data analysis demonstrates how various factors affect the thread consumption of the two thread types and how actually measured wastages are substantially lower than the buffer wastage that is included into the thread consumption worksheets. Measured wastages for TKT 120 and TKT 160 thread types throughout a range of operations are compared to the inbuilt buffer wastages, and the results show a continuous difference with the inbuilt buffer wastages being larger (Figure 4.1).

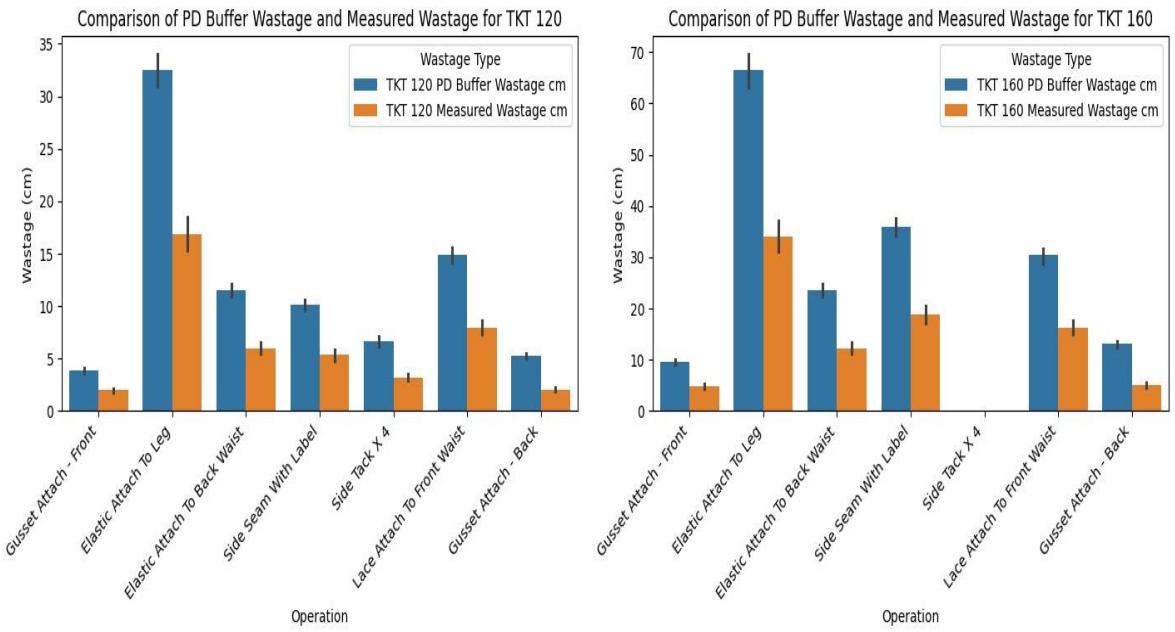


Figure 4.1 Buffer Wastage inbuilt by the Product Development vs Measured Wastage

The stitches per inch (SPI) or the stitch density has an immense influence on how much thread the machine operators use. Since more stitches are tightly packed into a certain seam length, a higher SPI is linked to higher thread utilization. More thread is required to travel the same distance with more stitches, demonstrating the significant impact of SPI on thread consumption (Figure 4.2).

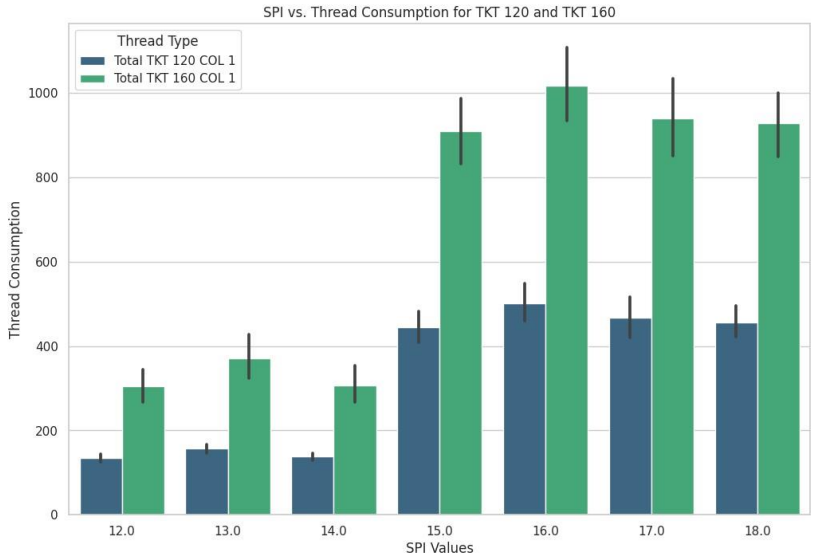


Figure 4.2 Influence of SPI on the thread consumption

The measurement of the stitched line or connection, known as the seam length, has a major impact on how much thread is used. As Figure 4.3 shows, a longer seam length usually necessitates using more thread to bind the fabric, increasing the thread consumption. The clusters that are visible in the scatter plot (Figure 4.3) that illustrates the correlation between seam length and thread consumption provides significant information about the thread that is used at different seam lengths in each of the seven different sewing operations. Every cluster comprises a collection of data points that exhibit a comparable thread consumption trend with respect to the relevant seam length.

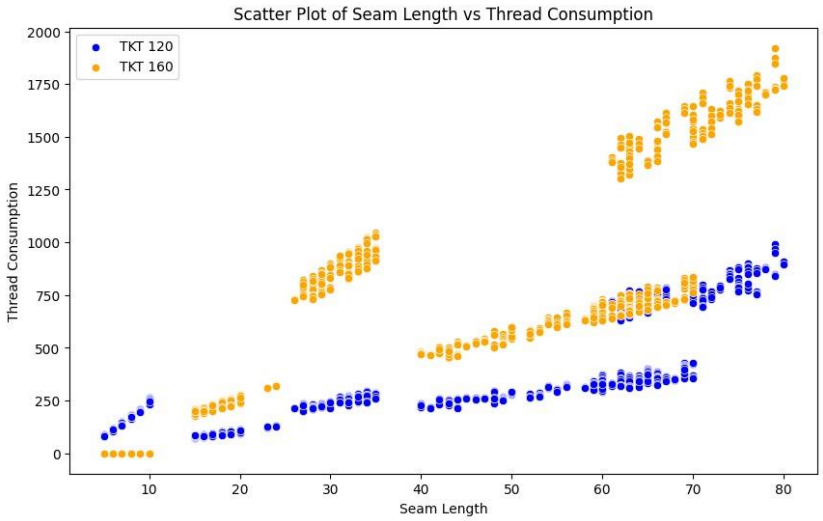


Figure 4.3 Influence of Seam Length on the thread consumption

The amount of thread used during the manufacturing of a garment is greatly influenced by seam thickness, which refers to the layers of cloth connected together by a seam. Often, more thread is needed to secure stitching on thicker seams, which is a sign of a more intricate sewing process. Different seam thicknesses affect tension and stress distribution, which in turn affects general thread consumption patterns. The link between seam thickness and thread consumption for TKT 120 and TKT 160 is depicted in the correlation matrices in Figure 4.4. Increased thread consumption in a variety of activities is correlated positively with higher seam thickness.

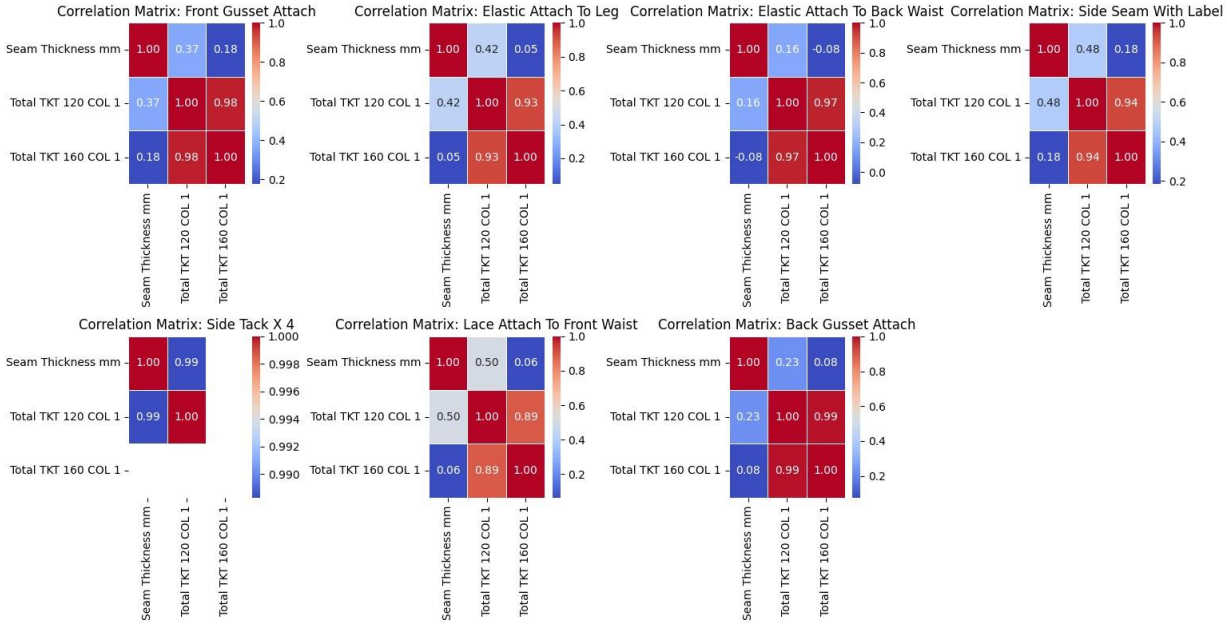


Figure 4.4 Correlations between Seam Thickness and thread consumption

The amount of thread that is expected to be wasted during the production process and contributes to the thread that is utilized but not included into the finished product is referred to as waste in thread consumption during the garment manufacturing process. Positive relationships are seen in Figure 4.5, which indicates that higher thread consumption is correlated with higher wastes.

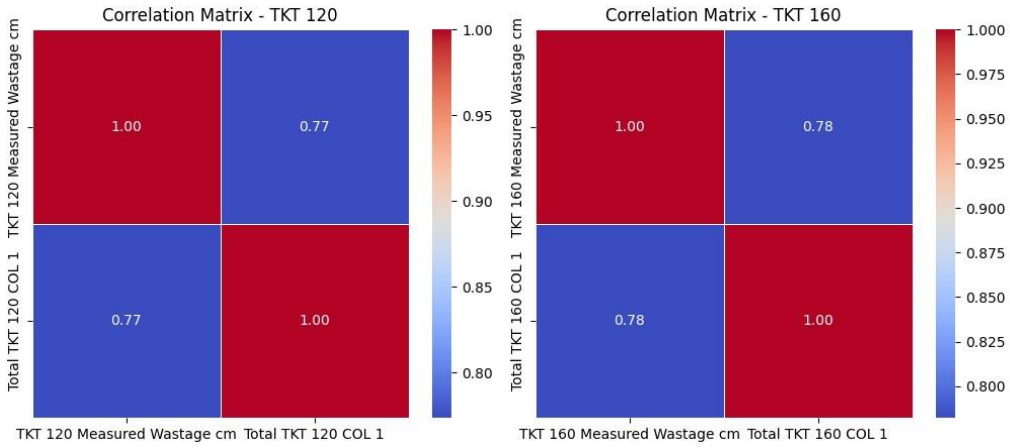


Figure 4.5 Correlation between Wastage and thread consumption

Thread consumption is mostly dependent on operator skill, which is a measure of sewing personnel's competence in the garment manufacturing process. Higher graded skilled operators (A+, A, B) show improved machine and operation knowledge, which results in optimal thread usage, decreased waste, and efficient sewing (Figure 4.6). The results point to a positive relationship between operator skill levels and optimal thread use, with a trend toward higher thread utilization being indicated by lower skill grade C.

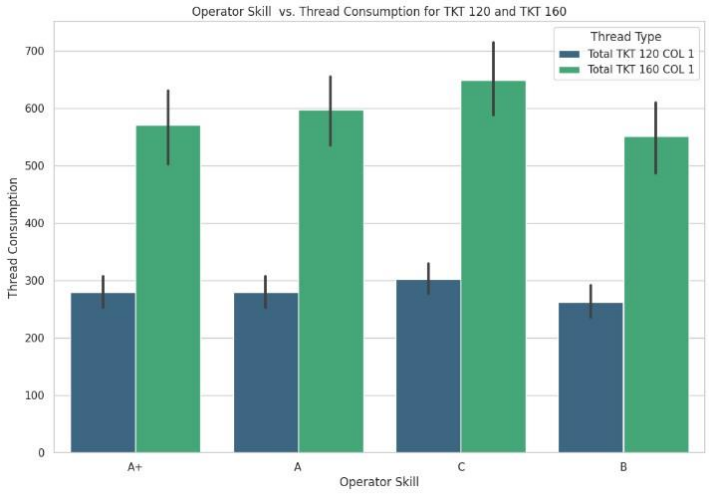


Figure 4.6 Influence of Operator Skill on the thread consumption

5. Evaluation Of Different Techniques Of Thread Consumption Prediction

The research shifted from using specific stitch types to estimate thread consumption to taking a more comprehensive approach. The emphasis changed to operation-wise predictions rather than isolated estimates for individual stitch types, offering a more thorough outlook on consumption for complete operations within a certain product category. Experimentation with multiple models, including Multivariate Linear Regression, Ensemble Models (Random Forest, Gradient Boosting, and XGBoost), and Multilayer Perceptron (Neural Network), was initiated due to having multiple predictions as the output rather than experimenting with single output prediction.

5.1 Multivariate Linear Regression

As an extension of the simple linear regression model, multivariate linear regression is used to jointly forecast the thread consumptions for TKT 120 and TKT 160. The model takes into account variables such as stitch length, density, seam length, etc. and assumes a linear relationship between various independent variables and thread consumptions. Tables 5.1 and 5.2 display the evaluation matrices (RMSE, MAE, MAPE, R-squared) for the wastage and consumption models.

Table 5.1 Evaluation matrices of Linear Regression Models - Wastage Model

Metrics for Wastage Model	***Before Feature Selection (with all the variables)***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	3.99	4.61	9.94	7.89	3.99	4.59	9.94	7.90
MAE	2.65	2.99	6.96	5.81	2.67	2.97	6.97	5.79
MAPE	57.97%	61.90%	60.28%	57.43%	58.33%	61.00%	60.28%	57.27%
R-Squared	0.6175	0.4078	0.5183	0.5648	0.6160	0.4124	0.5174	0.5633

Table 5.2 Evaluation matrices of Linear Regression Models - Consumption Model

Metrics for Consumption Model	***Before Feature Selection (with all the variables)***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	15.61	16.29	29.57	25.42	15.61	16.28	29.55	25.39
MAE	10.25	9.89	22.74	19.98	10.26	9.89	22.72	19.96
MAPE	4.79%	4.18%	4.12%	4.07%	4.78%	4.17%	4.12%	4.06%
R-Squared	0.9951	0.9946	0.9960	0.9966	0.9951	0.9946	0.9960	0.9966

After feature selection, the test dataset's error values dropped slightly, but the model's overall performance stayed the same. The MAPE values show a significant difference between the expected and actual values, particularly for the wastage model. On the other hand, the consumption model shows lower MAPE values, indicating a tighter match between the forecasts and the real observations. Figure 5.1 provides a visual representation of the actual vs. predicted values. It shows that TKT 120 has slightly bigger deviations than TKT 160, especially for high consumption rates.

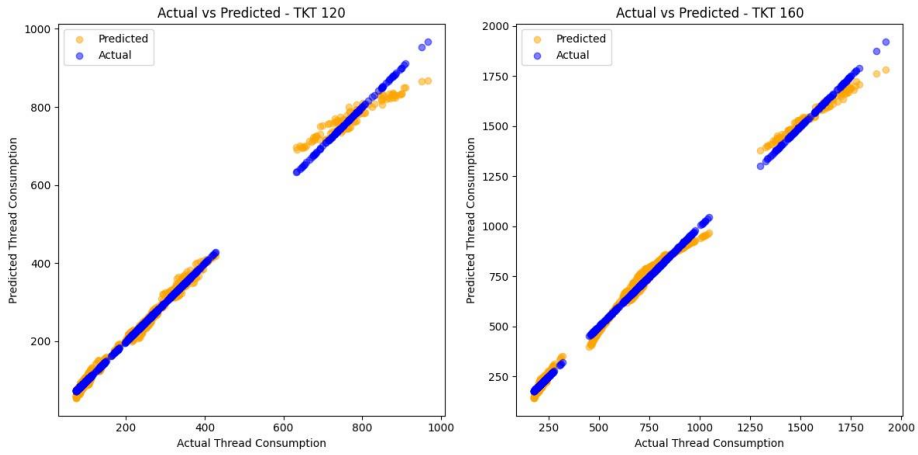


Figure 5.1 Actual vs Predicted values of Linear Regression model

5.2 Ensemble Models

A thorough summary of the ensemble models' evaluation matrices following feature selection is shown in Table 5.3. Based on the investigation, it can be shown that XGBoost performs better on the training dataset, as evidenced by its lowest error values. All ensemble models, including Random Forest and Gradient Boosting, share a common observation, though, which suggests that overfitting may have occurred. The discrepancy between very high error values on the test dataset and low error values on the training dataset makes this clear. This emphasizes how crucial it is to carefully evaluate model generalization for future applications in order to guarantee reliable performance in real-world situations.

Table 5.3 Consolidated Evaluation matrices of Ensemble model predictions for consumption

Metrics	***Random Forest***				***Gradient Boosting***				***XGBoost***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	3.08	7.91	7.27	11.64	5.66	7.29	12.9	13.91	2.21	6.5	5.46	10.12
MAE	2.01	4.51	4.8	8.13	4.00	4.68	9.16	10.03	1.24	3.91	3.26	6.89
MAPE%	0.70	1.43	0.68	1.24	1.68	1.65	1.39	1.71	0.54	1.28	0.52	1.15

5.3 Artificial Neural Network (ANN)

The Artificial Neural Network (ANN) model with 2 hidden layers underwent feature selection, excluding Operator Skill from training. Evaluation matrices were compared pre and post feature selection, with results detailed in Tables 5.4 and 5.5.

Table 5.4 Evaluation matrices of the ANN Wastage model pre and post feature selection

Metrics for Wastage Model	***With 2 hidden layers***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	***Test***	***Train***	***Test***	***Train***	***Test***	***Train***	***Test***
RMSE	2.23	4.61	5.77	7.96	1.95	2.98	3.89	4.79
MAE	2.33	2.98	6.27	5.86	1.23	2.64	3.44	4.56
MAPE	53.89%	58.36%	61.14%	58.17%	52.39%	57.71%	59.65%	59.95%
R-Squared	0.9995	0.9996	0.9989	0.9998	0.9995	0.9996	0.9989	0.9998

Table 5.5 Evaluation matrices of the ANN Consumption model pre and post feature selection

Metrics for Consumption Model	***With 2 hidden layers***				***After Feature Selection (with selected variables)***			
	TKT 120		***TKT 160***		***TKT 120***		***TKT 160***	
	Train	***Test***	***Train***	***Test***	***Train***	***Test***	***Train***	***Test***
RMSE	1.98	3.04	4.19	5.69	1.56	2.67	3.41	4.37

MAE	1.16	3.22	3.08	4.75	1.08	2.98	2.82	3.95
MAPE	1.15%	1.33%	1.99%	2.06%	1.12%	1.21%	1.76%	1.99%
R-Squared	0.9995	0.9996	0.9989	0.9998	0.9995	0.9998	0.9999	0.9997

With a smaller MAPE range of 50%–60%, feature selection significantly decreased error values for the Wastage Model (Table 5.4). A high R-Squared value indicates a well-fitting model. Similar post-feature selection improvements were shown by the Consumption Model (Table 5.5), which demonstrated lower error values and improved performance. Figure 5.2 displays similar results, with the predicted values being much closer to the actual values for both thread types TKT 120 and TKT 160.

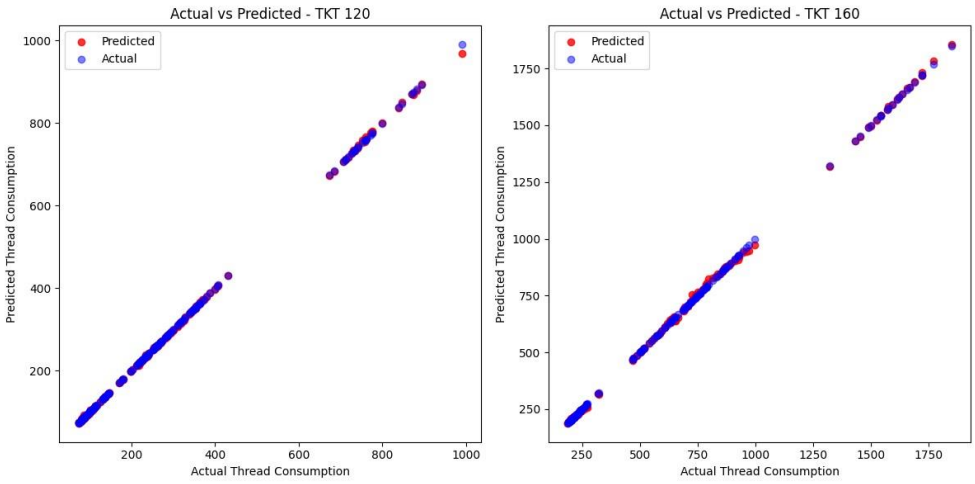


Figure 5.2 Actual vs Predicted values of ANN model

During ANN model training, the 'Early Stopping' approach was used to reduce overfitting nature. The outcomes demonstrated the ANN model's improved performance, outperforming other evaluated models in terms of predicting thread consumptions for two thread types in a single operation, taking buffer waste into account. This design offers a novel use of ANN models for thread consumption prediction, offering insightful information for streamlining processes in the apparel sector.

A comparison of different thread consumption prediction techniques designed by several researchers, against the proposed Artificial Neural Network (ANN) model is presented in Table 5.6. When MAPE is employed for evaluation, it becomes apparent that, while ANN models including the suggested model show lower MAPE values, geometrical and regression models typically have larger MAPE values. The table 5.6 further illustrates the reduced MAPE values, indicating that, the proposed ANN model performs better than other models in predicting thread consumptions for both TKT 120 and TKT 160.

Table 5.6 Comparison of different techniques of thread consumption prediction

	Geometrical Models			***Regression Models***			***Artificial Neural Network Models***					
Matric	***M1***	***M2***	***M3***	***M4***	***M5***	***M6***	***M7***	***M8***	***Proposed Model***			
									TKT 120		***TKT 160***	
									Train	***Test***	***Train***	***Test***
MAPE %	13.1	13.5	19.7	4.27	27.4	7.0	1.96	2.1	1.12	1.21	1.76	1.99

Notes: M1 - Model by (Ghosh & Chavhan, 2014), M2 - Model by (Jaouadi, et al., 2006), M3 - Model by (Chavan, et al., 2019), M4 - Model by (Jaouadi, et al., 2006), M5 - Model by (Abeysooriya & Wickramasinghe, 2014), M6 - Model by (Chavhan, et al., 2021), M7 – Model by (Jaouadi, et al., 2006) and M8 - Model by (Chavhan, et al., 2021)

6. Development Of A User Interface

Streamlit, a Python web application development tool, was used to create an interactive platform as the main focus of the study's user interface development. In particular, the proposed ANN models of the TKT 120 and TKT 160 thread types to be integrated together with the development environment setup and necessary libraries. These models were imported into the application using TensorFlow and Keras to create predictions based on user inputs received through interactive forms in the Streamlit interface. Figure 6.1 is a prototype of the user interface.

Figure 6.1 Proposed Interface for predicting thread consumption

Developing an intuitive and user-friendly interface was given top priority during the development process, and features were added and the layout was improved in response to user feedback and usability testing. Regardless of their level of technical expertise, users should be able to readily engage with the application and comprehend the predictions made by the machine learning models.

7. Conclusion

The aim of this study was to forecast the amount of sewing thread used in various sewing operations, with a specific focus on the thread types TKT 120 and TKT 160. The study meticulously worked through several critical phases of data preparation, such as feature engineering, cleaning, and addressing missing values, providing a strong basis for further studies. Important findings emerged from the exploratory data analysis, which suggested that estimated buffer wastages may have been overestimated in comparison to real wastage values. The study effectively handled its first objective, which was to determine the factors impacting thread consumption, including buffer wastage, stitches per inch (SPI), seam length, seam thickness and operator skill through the use of visualizations and statistical analyses

The proposed ANN models were found to be more accurate in prediction as compared to other techniques. The ANN network of thread type TKT120 shows mean absolute percentage errors 1.12% and 1.21% for training and testing whereas the ANN model of thread type TKT160 shows slightly higher mean absolute percentage errors 1.76% and 1.99% for training and testing. Other than the accuracy in prediction the proposed models are generalized, since the early stopping technique was implemented to mitigate overfitting nature of the model. For a specific apparel industry based on the varieties of fabric types and multiple thread types used, the ANN can be designed and trained accordingly.

Additionally, with ongoing training and historical data from a larger number of data samples, predictability can be raised to a maximum degree. Although the existing Artificial Neural Network (ANN) models show remarkable accuracy in anticipating thread consumption, further improvements can be made. The current models only consider thickness among fabric properties, relying on fundamental stitching settings and basic fabric properties. But other fabric characteristics, such as compression and shear, as well as the tension of the thread, might affect seam deformation and, in turn, thread consumption. These variables can be added to the models to help them better reflect the nuances of the sewing process. The desired enhancement entails going beyond simple prediction and toward optimization. This means designing models that predict thread consumption as well as seek to maximize thread utilization across a range of manufacturing products. Such advances would contribute for improving the estimation of thread consumption for a particular style, facilitating inventory management by reducing the amount of leftover stock and reducing significant costs to the organizations.

References

Abeysooriya, R. P. & Wickramasinghe, G. L. D., 2014. Regression model to predict thread consumption incorporating thread-tension constraint: study on lock-stitch 301 and chain-stitch 401.. *Fashion and Textiles* , Volume 1, pp. 1-8.

Abhishek, K., 2022. *Introduction to artificial intelligence.* [Online] Available at: https://www.red-gate.com/simple-talk/development/data-sciencedevelopment/introduction-to-artificial-intelligence/ [Accessed 25 12 2023].

Amaan Group, 2010. *Determining your sewing thread requirements.* s.l.:Amann Group.

American & Efrid Inc, 2007. *Technical bulletin estimating thread consumption..* s.l.:American & Efird Inc.

Baker, H. K., 2009. *Dividends and dividend policy.* Hoboken: John Wiley & Sons, Inc..

Barraza, N., Moro, S., Ferreyra, M. & De La Peña, A., 2018. Mutual information and sensitivity analysis for feature selection in customer targeting: A comparative study. *Journal of Information Science,* 45(1), pp. 53-67.

Chavan, M. V., Ghosh, S. & Naidu, M. R., 2019. An elliptical model for lockstitch 301 seam to estimate thread consumption.. *The Journal of the Textile Institute,* Volume 110(12), pp. 1740-1746.

Chavhan, M. V., Naidu, M. R. & Jamakhandi , H., 2021. Artificial neural network and regression models for prediction of sewing thread consumption for multilayered fabric assembly at lockstitch 301 seam. *Research Journal of Textile and Apparel,* 26(4), pp. 343-358.

Coats Digital, 2023. *Coats SeamWorks.* [Online] Available at: https://www.coats.com/en/solutions/coats-seamworks [Accessed 10 08 2023].

DOĞAN, S. & PAMUK, O., 2014. Calculating the amount of sewing thread consumption for different types of fabrics and stitch types. *TEKSTİL ve KONFEKSİYON,* 24(3).

Emjay International (Pvt) Ltd., 2023. *About.* [Online] Available at: http://www.emjayi.com/

Ghosh, S. & Chavhan, M. V., 2014. A geometrical model of stitch length for lockstitch seam. *Indian Journal of Fibre and Textile Research,,* 39(2), pp. 153-156.

Jaouachi , B. & Khedher , F., 2015. Evaluation of Sewed Thread Consumption of Jean Trousers Using Neural Network and Regression Methods. *FIBRES & TEXTILES in Eastern Europe,* 3(111), pp. 91-96.

Jaouadi, M., Msahli , S., Babay , A. & Zitouni , B., 2006. Analysis of the modelling methodologies for predicting the sewing thread consumption. *International Journal of Clothing Science and Technology,* Volume 18, pp. 7-18.

Khedher, F. & Jaouachi, B., 2015. Waste factor evaluation using theoretical and experimental jean pants consumptions. *The Journal of The Textile Institute,* 106(4), p. 402–408.

Mariem, B. et al., 2020. A Study of the Consumption of Sewing Threads for Women's Underwear: Bras and Panties. *AUTEX Research Journal,* 20(3), pp. 299-311.

Natekin, A. & Knoll, A., 2013. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics,* Volume 7.

Noriega, L., 2005. *Multilayer Perceptron Tutorial,* s.l.: s.n.

Olive, D., 2017. *Linear regression.* s.l.:Springer.

Rengasamy, R. S. & Samuel, W. D., 2011. Effect of thread structure on tension peaks during lock stitch sewing. *AUTEX Research Journal,* 11(1), pp. 1-5.

Ribeiro, J., Lima, R., Eckhardt, T. & Paiv, S., 2021. Robotic process automation and artificial intelligence in industry 4.0. *Procedia computer science,* 181(1), pp. 51-58.

Ukponmwan, J. O., Mukhopadhyay, A. & Chatterjee, K. N., 2000. *Textile Progress.* s.l.:The Textile Institute.

Vasiliev, V. A., Velmakina, Y. V. & Mayborodin, A. B., 2019. Using artificial neural networks when integrating the requirements of standards for management systems in QMS. *IOP Conference Series: Materials Science and Engineering,* 666(1), p. 012058.

Yeşilpınar, S. & Alkiraz, F., 2005. Kumaş kalınlığının dikiş iplik giderine etkisinin incelenmesi. *The Journal of Textiles and Engineer,* 12(59-60), pp. 29-34.