SEMARAK ILMU PUBLISHING
202103268166(003316878-P)

Semarak Engineering Journal

Journal homepage:
https://semarakilmu.my/index.php/sej/index
ISSN: 3036-0145

Production Capacity Forecasting Model: A Data Approach in the Apparel Industry

Song Thanh Quynh Le[1], Thanh Nhi Huynh Thi[1], Thanh Nhan Phan[1], Mai Ha Phan[2,*]

[1] Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology (HCMUT), VNU-HCM, Ho Chi Minh City, Vietnam
[2] School of Industrial Management, Ho Chi Minh City University of Technology (HCMUT), VNU-HCM, Ho Chi Minh City, Vietnam

ARTICLE INFO

Article history:
Received 22 May 2026
Received in revised form 13 August 2026
Accepted 27 August 2026
Available online 7 September 2026

Keywords:
Production capacity; clothing industry; customer order; forecasting model; machine learning

ABSTRACT

In the clothing industry, a sector facing increasingly stringent demands regarding productivity, costs, and delivery schedules, the application of machine learning to forecast production capacity has become an essential management requirement. This study focuses on developing a production capacity forecasting model by analyzing the interplay between clothing order characteristics and sewing line parameters. The Naive Bayes model provides a probabilistic framework for classifying expected production capacity into distinct levels; consequently, the model's output offers a clear basis for planners to assess capacity-related risks and make informed decisions regarding resource allocation and line balancing. Specifically, the forecasting results can support decisions concerning workforce requirements, production scheduling, sewing line assignments, and delivery planning. The research also demonstrates the potential of machine learning as a decision-support tool, transforming historical production data into actionable insights to inform practical plans. These findings will enable fashion enterprises to optimize resources, meet rigorous customer demands, and mitigate risks throughout the production process.

1. Introduction

In an increasingly developed and competitive apparel industry, forecasting and optimizing production capacity have become crucial factors in determining enterprise success [1]. In addition, many constraints on sustainable development policies in the supply chain have placed significant pressure on stakeholders, especially product manufacturers [2]. Apparel enterprises not only face high customer expectations regarding delivery time and product quality but also must optimize resources to minimize costs and enhance production efficiency. This presents a significant challenge for planning and managing production capacity, given the distinct characteristics of orders across markets [3]. In addition, the clothing industry in many countries is shifting from the Original Equipment Manufacturer (OEM) model to the Original Design Manufacturer (ODM) model, in which

* *Corresponding author.*
E-mail address: ptmaiha@hcmut.edu.vn

https://doi.org/10.37934/sej.16.1.3448

businesses assume greater responsibility for the entire product design and development process, including production and supply chain management. This requires businesses to enhance their forecasting and production planning capabilities to ensure flexibility and operational efficiency. However, many clothing enterprises still use manual or experience-based methods to predict production capacity. These methods are not only inaccurate but also have numerous consequences, including inefficient resource allocation, wasted time, and higher production costs. Meanwhile, data-driven forecasting models have proven highly effective at enhancing planning capabilities, particularly in modern manufacturing industries [4]. Applying production capacity forecasting models not only helps enterprises meet customer demands promptly but also enhances their market competitiveness.

In production activities, in general, and the clothing industry in particular, the productivity of the production line is influenced by numerous factors that interact with and affect one another simultaneously, especially order characteristics. This paper presents a production capacity forecasting model based on an analysis of the interaction between order characteristics and production line specifications in clothing enterprises, which is not only timely in a practical context but also holds significant scientific and practical implications. Researching and developing an effective forecasting model will enable garment enterprises to optimize resources, meet stringent customer requirements, and mitigate production risks.

2. Background

2.1 Definition of Production Capacity

Production capacity refers to the maximum output a factory, production line, or workstation can produce within a given period under normal operating conditions [5]. In reality, the production line experiences frequent fluctuations, preventing its capacity from reaching its maximum. In the clothing industry, production capacity is expressed in three forms: theoretical, effective, and realized capacity [6].

- *Theoretical capacity* is the maximum level of performance that can be achieved under perfect conditions, with no downtime, no product defects, and no interruptions.
- *Effective capacity* is the expected level of capacity that a sewing line can achieve in everyday production practice, taking into account factors such as downtime, product defects, or line changes.
- *Realized capacity* is the actual capacity level after subtracting unplanned constraints such as machine maintenance, worker leave, or delays in the supply of raw materials.

2.2 Factors affecting production capacity in clothing production

In production management, identifying the determinants of production capacity is essential for accurate capacity forecasting and effective risk mitigation, thereby ensuring the achievement of established productivity targets. In the clothing industry, sewing line capacity is influenced by numerous factors that interact simultaneously and mutually [7]. To comprehensively evaluate these factors, the "5M + 1E" model is employed as an effective analytical tool, enabling the identification of causes and the provision of appropriate improvement solutions. The “5M + 1E” model is a popular root cause analysis method, often applied in quality management systems and industrial production. This model originates from the Fishbone Diagram tool, developed by Kaoru Ishikawa, a Japanese quality management expert, in the 1960s [8].

a) Man factor

In the clothing industry, workers are considered the key factor, directly determining the operational capacity of the sewing production line. They are responsible for maintaining stable productivity and ensuring that products meet technical standards [9]. The productivity of the sewing line depends almost entirely on the workers' skill and experience. Additionally, the stability of the sewing line also depends on the operator's skills, their ability to control seam quality, and their capacity to respond flexibly when errors occur on the product [10].

In clothing enterprises, workers' skill levels are used as the standard for ranking skills. The higher the skill level of workers, the more capable they are of undertaking key workstations on the production line, which require higher operating skills and precision to maintain stable production capacity [11]. In managing the production line, the sewing line manager often bases their decisions on workers' skill levels and practical experience to assign suitable workers to workstations, ensuring the sewing line meets delivery times and fulfills the technical requirements of each order. Sewing lines with varying numbers of workers, each with different skill levels, will have different production capacities.

b) Machine factor

The machinery and equipment in the production line are considered key technical components that directly determine production capacity [12]. Machinery not only plays a supporting role in sewing operations but also significantly affects the speed, stability, and quality of the output products. The ability to invest in and apply modern machinery and technology enables the sewing line to perform complex operations at higher speeds while eliminating many redundant manual operations. This helps shorten processing times, limit waiting for semi-finished products between stages, and thereby increase overall capacity [13].

However, the efficiency of machinery also depends on regular maintenance and repair. Reasonable control of equipment status helps maintain stable operation and minimize technical problems and downtime [14]. On the contrary, when machines break down frequently, processes are interrupted, and actual capacity will decrease significantly from the planned level.

c) Material factor

In a clothing production line, raw materials are crucial, as they directly affect the processing capacity and operational efficiency of the sewing line. Raw materials not only reflect the complexity of the product but also affect operational speed, error rate, process stability, and, ultimately, actual capacity [15]. The characteristics of raw materials clearly affect the ease or difficulty of the sewing tasks [16]. Several material-related factors directly affect the sewing line's production capacity, including fabric physical properties, color, and the number of colors in the clothing order.

- *The physical properties of fabric*: Fabric weight is a crucial factor that directly impacts the productivity of workers in the sewing line. For heavy fabrics, the pulling force during sewing is greater, making the needle tip more prone to breakage or causing the seam to deviate if the worker operates too quickly [17]. Therefore, it is often necessary to reduce the sewing speed to ensure safety and accuracy. For fabrics that are too thin and prone to shifting during sewing, workers must maintain a slow, steady rhythm to keep the seam stable and meet technical requirements. In addition, the fabric's elasticity also affects the line's production capacity. Highly elastic fabrics often require specialized sewing techniques to ensure that seams remain intact during processing and in use. When working with elastic fabrics, sewing workers must maintain a consistent pulling force and adjust the machine speed appropriately to prevent the fabric from stretching or distorting and the thread from coming loose. This means that the operating speed is often slower than when sewing non-elastic fabrics. Additionally, elastic fabrics require specialized equipment and needles to maintain their elasticity in the finished product. If not equipped with the right machinery, the seams are very likely to break or leak, increasing the rate of defective products and requiring repair steps, which directly affect the line's actual capacity.

- *The fabric's colo*r: customers often have strict requirements for color uniformity across the entire product. This forces businesses to use cut-and-match details from the same fabric or dye batch to avoid shade differences. In the event of a significant shade difference, the sewing line must separate the batch, re-arrange the order of use, or even stop to re-sort the items. Additionally, each additional color in an order requires preparation operations, such as changing threads, adjusting cutters, and setting up the appropriate machines or tools [18]. These operations increase preparation time, generate many line stops, and reduce the total actual production time. The more colors an order has, the longer the changeover time and the greater the preparation work, reducing overall production capacity.

d) Method factor

The production method reflects the organization of the sewing process, the level of technical complexity, the number of stages, and the ratio of manual to semi-automatic operations [17]. In actual production, the characteristics of the sewing process often vary significantly depending on the product's nature. The more stages are automated, the lower the manual operation ratio, the shorter the processing time, and the more stable the quality. Another important factor is whether the enterprise has established and implemented a Standard operating procedure (SOP). SOP helps standardize operations at each stage, thereby reducing errors caused by skill differences among workers. Combined with effective line balancing, it will help optimize labor distribution across stages, reduce bottlenecks, and increase actual productivity. According to [19], applying the appropriate line-balancing method can increase productivity by 15% to 20% compared to unscientific layout processes.

e) Measurement factor

Measurement is a group of factors that reflect quality inspection and monitoring activities throughout the production process. Quality control indicators not only ensure that products meet customers' technical requirements but also directly affect actual production speed and productivity [20]. In a clothing manufacturing line, quality control activities depend on customer requirements and directly impact production capacity. This is evident in the fact that different quality standards require different inspection frequencies, control levels, and evaluation methods. The higher the quality requirements, the more time is spent on inspection and error correction, thereby altering the takt time between workstations. Without proper organization, inspection points can become bottlenecks, causing congestion and imbalances in the production line. Conversely, a well-established quality control system will maintain a stable production flow, thereby increasing productivity and overall line efficiency.

f) Environment

The working environment is the last factor, but it plays a crucial role in maintaining stable labor performance among sewing workers. Environmental conditions in the factory directly affect health, physical condition, concentration ability, and operating speed [21]. Several environmental factors affect the sewing line's capacity, including temperature, humidity, noise, ventilation, and lighting [22]. In sewing production, most stages require workers to observe sewing details, seams, and fabric edges with high precision. If lighting conditions are insufficient or substandard, workers are likely to experience eye fatigue and strain, leading to slower operations, misaligned seams, and technical errors. In particular, with dark fabrics, hidden patterns, or products that require sewing small details, weak lighting will make it almost impossible to detect errors in a timely manner. This leads to increased error-correction

2.3 Data mining approach in production management

The emergence of data mining underscores the importance of efficient access to and use of information in production management. It can be described as a method for uncovering the "meaning" in data, enabling users to gain the insights needed to make informed decisions [23]. Data mining is particularly effective when dealing with large, complex datasets that contain many variables and exhibit non-linear relationships. It helps predict behavior and outcomes, and uncover hidden patterns and associations, to gain deeper insights.

Data mining techniques can be categorized into two types of learning: supervised and unsupervised [24]. In supervised learning, classification is the preferred method in the clothing industry, especially with Naive Bayes, due to its ability to analyze complex patterns, optimize production processes, and enhance quality control.

In data mining, Naive Bayes is a probabilistic classifier based on Bayes' Theorem that assumes independence among predictor variables [25]. While this “naïve” assumption might not always hold true in real-world scenarios, Naive Bayes often performs surprisingly well in practice, especially with high-dimensional data. According to [26], Naive Bayes has several advantages, including simplicity, speed, and high accuracy. However, it also has some disadvantages related to its “Naïve” nature and limited structure. Theorem is stated as follows in *Eq.1:*

$$(Y = y_k|X) = \frac{P(X|Y = y_k) \times P(Y=y_k)}{P(X)} \tag{1}$$

Where:
– $P(Y = y_k|X)$ is the posterior probability, the probability of the class y_k given the predictor attributes $X = (X_1, X_2 \dots\dots, X_m)$.
– $P(X|Y = y_k)$ is the likelihood, the probability of observing attributes X given class y_k.
– $P\ (Y = y_K)$ is the prior probability of class y_k.
– $P(X)$ is the probability of observing attributes X (evidence). Since $P(X)$ is constant for all classes, it is usually ignored for classification purposes, and predictions are made based on the numerator $P(X|Y = y_k) \times P(Y = y_k)$.

According to [27], the production capacity in most cases cannot be measured by a crisp number; its proper identification requires a distribution of values and a probability statement attached to it. Therefore, applying the Naive Bayes algorithm in this study to predict production capacity based on influencing factors from "5M and 1E" is theoretically reasonable, as it aligns with the probabilistic interpretation of production capacity and supports risk-based decision-making in production planning.

3. Predicting The Production Capacity Method – A Case Study

3.1 Production capacity management process at an apparel company

Assessing production capacity is a crucial step in clothing production management, as it directly impacts the ability to accept orders, plan, and organize production effectively. At X Clothing Company, the current production capacity assessment process is carried out in four steps. However, this process still has many limitations, particularly in linking order characteristics to the sewing line's execution capacity, which results in low forecast accuracy.

Step 1 - Collecting the order's information: Order information collected from customers by the sales and merchandising departments includes: order type (CM/FOB), quantity, model, specifications, size chart, delivery time, technical requirements, raw materials, and other relevant details. This forms the basis for assessing the order's production capacity. The order management system fully supports this information and is easily accessible, allowing access to historical order data. However, this process currently lacks tools to quantify the complexity/technical level of each order, leading to preliminary assessments that do not accurately reflect the actual difficulty.

Step 2 - Resource analysis: Resources on the sewing line are analyzed based on fundamental factors, including the number of workers, shifts, the condition and quantity of machinery, and the availability of materials. This analysis is primarily based on the sewing line supervisor's experience. The raw data lacks a systematic quantitative approach to assessing skill levels and the suitability of resources for each product type; assessments remain subjective.

Step 3 - Assessing actual production capacity: X company's actual production capacity is determined primarily by the average final output of each production line or the factory as a whole, using historical data as a basis. This method provides a general estimate of production capacity but does not differentiate between the characteristics of individual orders, such as processing difficulty, materials, or product design.

Step 4 - Compare production capacity with order demand and make decisions: After estimating production capacity, company X compares it with the output required for each order. The comparison results serve as the basis for the company to decide whether to accept or reject the order, or to re-allocate workforce, time, and equipment to meet the deadline. An explicit limitation is the company's lack of clear quantitative tools for assessing capacity relative to the actual needs of each order. Decisions are primarily based on managers' experience and subjective judgment, which can lead to errors in resource allocation.

Analysis of the current production capacity assessment process at X Company reveals that the current method is not closely aligned with the specific characteristics of each order. Capacity assessment and decision-making are primarily based on historical data and practical experience, which can lead to an inaccurate reflection of actual production capacity in the context of increasingly diverse and flexible orders, particularly in terms of order type, material, and the complexity of production tasks. Given this situation, there is an urgent need to develop a production capacity forecasting model based on the characteristics of each order combined with sewing line technical characteristics, aiming to help the company achieve the following objectives:

- Accurately estimate the ability to fulfill new orders, ensuring that decisions to accept or reject orders are based on actual data.
- Reduce the risk of accepting orders beyond actual capacity, avoiding delivery delays or fluctuating productivity.

3.2 Developing the production capacity forecasting model

By developing forecasting models that combine clothing order characteristics with manufacturing line characteristics, fashion businesses can establish a basis for adjusting, controlling, and optimizing production factors, thereby improving on-time delivery rates, maintaining stable production efficiency, and enhancing competitiveness in an increasingly dynamic market. The methodology of the production capacity forecasting model is shown in Fig. 1.

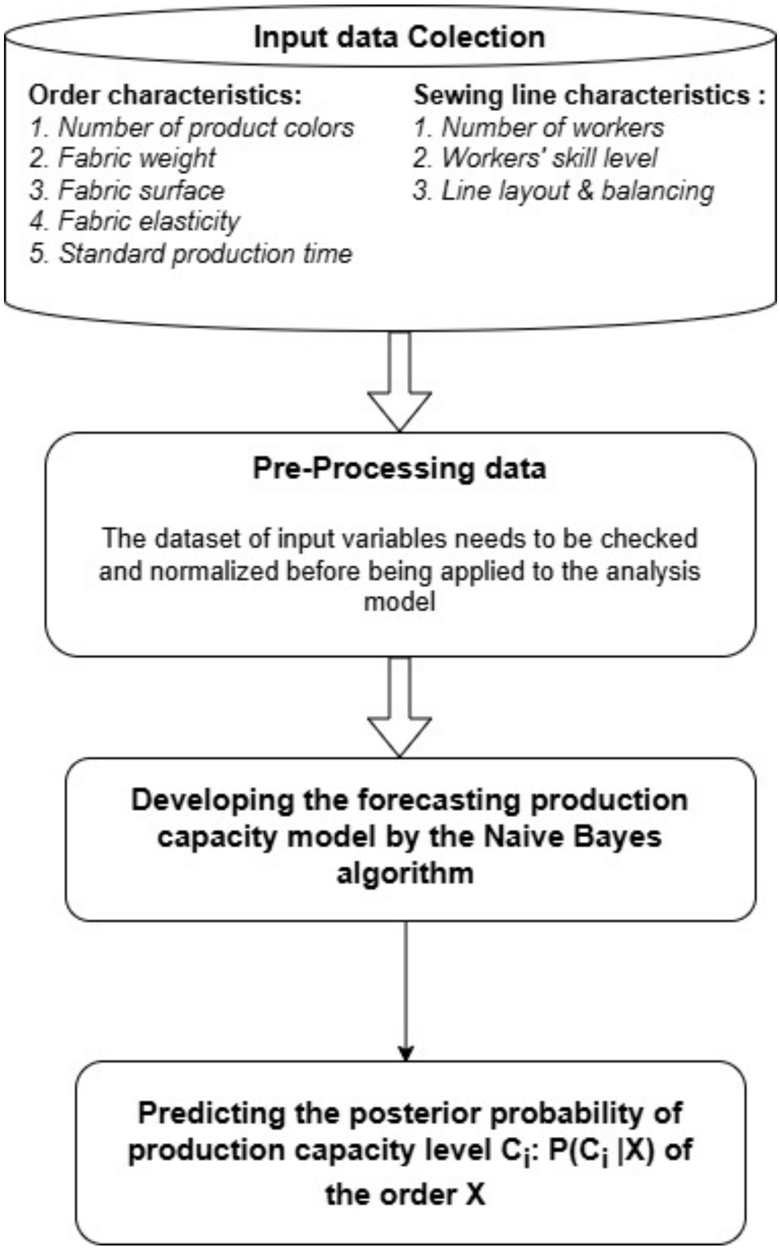

Fig. 1. The methodology for the production capacity forecasting model

Step 1 - Collecting the input data

In this step, the input variables of the forecasting model will be identified. These order-related factors fluctuate and, when combined with the sewing line's production characteristics, affect production capacity. Based on the literature review in Section 2 on "5M + 1E," the descriptive variables data will be separated into two groups*: order characteristics* and *production line characteristics*. In this model, the group of factors related to order characteristics primarily inherits components from the "Material" and "Method" categories in the 5M and 1E models, reflecting factors such as product type, technical requirements, complexity, and order size. Additionally, the group of factors related to the production line includes components from "Man," "Machine," "Method," and a portion of "Measurement," representing the internal operational capabilities of the production line, such as the number and skill level of workers, line organization, working conditions, and control systems. "Environment" factors remain almost constant across production lines within the same factory.

The transition from the 5M and 1E analytical framework to the new grouping method does not change the nature of the factors; it aims to streamline the identification of data sources and storage locations, thereby increasing the feasibility of data collection. The mapping of the 5M1E model factors into two groups is described in Table 1.

Table 1
Mapping of the 5M1E model to two groups of factors

5M+1E Model	***Order characteristics***	***Production line characteristics***
Man factors		Workers' skill level
Machine factors		Line layout
Material factors	Fabric weight, Fabric surface, Fabric elasticity	
Method factors	Number of colors, Standard production time	Number of workers
Measurement factors		Line Balancing
Environment factors	x	x

1. Descriptive variables

a) Order characteristics

Order characteristics are crucial factors that directly influence the operation list, worker assignment, and machinery layout, thereby impacting sewing line capacity [28]. These characteristics reflect the order's volume, complexity, and technical specifications, thereby determining the difficulty of operations and the time required to complete the product. The main input variables used to model and forecast production capacity, including:

- *Number of product colors*: is a quantitative variable that affects machine setup time, thread change time, synchronized control of raw materials and accessories, and the speed of worker operations.
- *Fabric weight (GSM)*: a quantitative variable that directly affects the operation speed; fabrics that are too thick or too thin will reduce sewing efficiency.
- *Fabric surface*: considered a qualitative variable on an ordinal scale. This characteristic directly affects the ability to control fabric edges and the stability of seams during production.
- *The elasticity of a fabric* directly affects the speed of operation and the seam error rate in production. The greater the fabric's elasticity, the more frequently workers must adjust and secure the fabric edges during production to ensure product quality.
- *Standard production time*: is the time allocated to complete a product, reflecting its complexity and serving as a basis for estimating production line productivity.

b) Sewing line characteristics

The group of line characteristics includes factors directly related to the sewing line's productivity and operational efficiency. These characteristics determine the ability to complete products, the stability of the line throughout the production process, and the allocation of workers and the balance of processes. The main input variables of this group include:

- *Number of workers:* a quantitative factor reflecting the total number of workers on a sewing line. The number of workers directly affects the volume of finished products per day and the efficiency of task allocation.
- *Worker skill level:* this reflects the professional competence of workers on the sewing line, classified by the average skill level of the line, from skill level 3 to 5. This variable reflects the line's operational capacity and the speed at which products are completed.

- *Line Layout & Balancing:* line balancing not only affects average productivity but also determines the line's stability during the production process.

2. Class variable

The actual production capacity of the sewing line should be compared with the delivery time constraints and the customer's required productivity levels to provide a basis for adjusting production plans and allocating resources rationally. Based on this comparison, production capacity can be divided into three levels:

- *Under capacity* occurs when the maximum output achievable within the allowed time frame is lower than the required output, which implies a risk of delays and may require interventions such as overtime, additional labor, or schedule adjustments.
- *Just sufficient capacity* is when actual capacity is approximately at the required level, ensuring order completion but carrying the risk of unforeseen fluctuations.
- *Excess production capacity* occurs when actual capacity exceeds customer demand, enabling the business to flexibly accept additional orders or optimize costs by reallocating resources.

Categorizing competence into these three states not only helps quantify responsiveness but also provides a foundation for probabilistic classification models in forecasting and decision support.

Step 2 - Pre-processing data

After data collection, the dataset of input variables needs to be checked and normalized before being applied to the analysis model. The purpose of this step is to eliminate inaccurate, incomplete, misformatted, duplicated, or irrelevant information, thereby minimizing bias in the analysis results and improving the accuracy of the production capacity forecasting model. The input data for the forecasting model contains both qualitative and quantitative variables, so data normalization is necessary. Quantitative variables need to be mapped to ordinal scales in order of their increasing influence on production capacity. ***Table 2*** presents the descriptive statistics for five quantitative variables in the forecasting model. During this step, five technical and quality management specialists from the apparel firm are selected to serve on an expert panel.

Table 2
The descriptive statistics of five quantitative variables

Variables	Min	Max	Mean	Median	Standard deviation
Number of product colors	1	4	2.19	2	1.10
Fabric weight (g/m^2)	100	179	138.43	136	20.58
Standard production time (minutes/pcs)	15	34.02	22.30	21.77	4.53
Number of workers	20	40	31.58	33	5.19
Balancing & Layout Efficiency	0.62	1.17	0.89	0.88	0.11

Two of them work in the manufacturing department, one in the planning department, and two in the quality assurance department. The conversion of quantitative variables into ordinal variables is based on descriptive statistics of the data, combined with expert panel checks for reasonableness and consistency. This process helps mitigate the impact of measurement errors and outliers while better reflecting the hierarchical nature of input variables in line with actual production practices. Table 3 then uses categorical data to illustrate the categorization of quantitative variables.

Table 3
The categorization of input variables

Input variables	Categorizational level
Number of colors	Little ($1 \leq x \leq 2$); Medium ($2 < x \leq 4$); Large ($x > 4$).
Fabric weight (GSM)	Light ($100 \leq GSM \leq 127$); Medium ($127 < GSM \leq 154$); Heavy ($GSM > 154$).
Fabric surface	Slippery; Medium; Rough.
Fabric elasticity	Low; Medium; High.
Standard production time	Low ($15 \leq minutes \leq 21.5$); Medium ($21.5 < minutes \leq 28$); High ($> 28\,minutes$).
Number of workers	Little ($20 \leq y \leq 27$); Medium ($27 < y \leq 34$); Large ($y > 34$).
Workers' skill level	Level 3; Level 4; Level 5.
Line Layout & Balancing Efficiency	Low ($0.6 \leq z \leq 0.8$); Medium ($0.8 < z \leq 0.95$); High (> 0.95).

Step 3 - Application of Naive Bayes for predicting production capacity

From the result of production capacity in 463 historical orders, each production capacity level $P(C_i)$ probability is estimated as follows:

$$P(production\ capacity = "Under\ capacity")$$
$$= \frac{The\ number\ of\ under\ capacity\ production\ order}{Total\ oders} = \frac{190}{463} = 0.41$$
$$P(production\ capacity = "Sufficient\ capacity")$$
$$= \frac{The\ number\ of\ sufficient\ capacity\ production\ order}{Total\ oders} = \frac{173}{463} = 0.37$$
$$P(production\ capacity = "Excess\ capacity")$$
$$= \frac{The\ number\ of\ excess\ capacity\ production\ order}{Total\ oders} = \frac{100}{463} = 0.21$$

The prior probability values of eight descriptive attributes $P(x_k|C_i)$ are calculated, as shown in ***Table 4***.

Table 4
The prior probability values of each descriptive attribute

Attribute	***Attribute value***	$P(x_k \mid C_i)$		
		Under capacity	Sufficient Capacity	Excess capacity
Number of colors	Little	108 ***(0.57)***	96 ***(0.55)***	94 ***(0.94)***
	Medium	46 ***(0.24)***	31 ***(0.18)***	5 ***(0.05)***
	Large	36 ***(0.19)***	46 ***(0.27)***	1 ***(0.01)***
Fabric weight	Light	61 ***(0.32)***	37 ***(0.22)***	20 ***(0.20)***
	Medium	84 ***(0.44)***	89 ***(0.51)***	78 ***(0.78)***
	Heavy	45 ***(0.24)***	47 ***(0.27)***	2 ***(0.02)***
Fabric surface	Slippery	75 ***(0.40)***	40 ***(0.23)***	4 ***(0.04)***
	Medium	59 ***(0.31)***	68 ***(0.39)***	64 ***(0.64)***
	Rough	56 ***(0.29)***	65 ***(0.38)***	32 ***(0.32)***
Fabric elasticity	Low	69 ***(0.36)***	77 ***(0.44)***	94 ***(0.94)***
	Medium	89 ***(0.47)***	83 ***(0.48)***	5 ***(0.05)***
	High	32 ***(0.17)***	13 ***(0.08)***	1 ***(0.01)***
Standard time	Low	52 ***(0.27)***	56 ***(0.32)***	93 ***(0.93)***
	Medium	94 ***(0.50)***	91 ***(0.53)***	5 ***(0.05)***
	High	44 ***(0.23)***	26 ***(0.15)***	2 ***(0.02)***
Number of workers	Little	53 ***(0.28)***	36 ***(0.21)***	1 ***(0.01)***
	Medium	84 ***(0.44)***	77 ***(0.44)***	36 ***(0.36)***
	Large	53 ***(0.28)***	60 ***(0.35)***	63 ***(0.63)***
Skill level	Level 3	154 ***(0.81)***	130 ***(0.75)***	36 ***(0.36)***
	Level 4	33 ***(0.17)***	39 ***(0.23)***	63 ***(0.63)***
	Level 5	3 ***(0.02)***	4 ***(0.02)***	1 ***(0.01)***
Line efficiency	Low	58 ***(0.31)***	39 ***(0.23)***	1 ***(0.01)***
	Medium	131 ***(0.68)***	131 ***(0.76)***	3 ***(0.03)***
	High	1 ***(0.01)***	3 ***(0.02)***	96 ***(0.96)***

4. Results and Discussion

Table 5 summarizes the performance of the Naïve Bayesian model. The prediction model's accuracy is about 68.3%. However, relying solely on the accuracy indicator to evaluate a model is insufficient. Accuracy reflects the percentage of correct predictions across the entire dataset, but does not clearly show the model's effectiveness on specific classes, especially in multi-class classification problems. Detailed analysis by class reveals significant differences in predictive performance.

Specifically, the model performed very well for the "Excess capacity" class, with a true positive rate of 0.94, a precision rate of 0.98, and a very low false positive rate (0.008). This indicates that the model is highly accurate at identifying cases of excess capacity and rarely confuses them with other classes.

Conversely, for the “Under capacity” class, performance is only average with a true positive rate of 0.65 and a precision of 0.64. The relatively high false positive rate (0.26) indicates that the model still makes significant mistakes when distinguishing this class from others. Notably, the “Sufficient capacity” class has the lowest performance, with both the true positive rate and precision reaching only 0.58. This suggests that the model has difficulty distinguishing the sufficient-capacity state from the other two states. These results suggest that evaluating forecasting models should not be limited to accuracy alone but should also consider other metrics, such as true positive rate (recall), precision, and false positive rate. These metrics provide a more comprehensive view of the model's classification capabilities across classes, helping to identify potential weaknesses that accuracy may not reflect.

Table 5
Performance of the Naïve Bayesian model

Model		Naïve Bayesian	
Correctly classification instances		68.3%	
Incorrectly classification instances		31.7%	
Detailed accuracy by class			
Class	True positive rate	False positive rate	Precision
Under capacity	0.65	0.26	0.64
Sufficient capacity	0.58	0.26	0.58
Excess capacity	0.94	0.008	0.98

Production capacity forecasting models are valuable tools that support clothing companies and operations management departments in more accurately planning their capacity to meet customer orders, helping identify potential schedule risks in the production process before they occur and providing a foundation for selecting the appropriate production line configuration to meet schedule requirements.

Based on the prior probability values of each descriptive attribute, when a new order X is placed with the following attributes, including the number of colors in the order is medium, the fabric material used for the order is light-weight, the fabric surface is slippery, and the elasticity is medium. In addition, after analyzing the technical production process, the production standard time is medium. If this order is produced into a sewing line with the following characteristics: a low number of workers on the line, an average worker skill level of level 3, and a generally medium line balancing efficiency. The attribute vector will be:

X = (Number of colors is medium, fabric weight is light, fabric surface is slippery, fabric elasticity is medium, standard time is medium, number of workers is little, average skill levels is level 3, and line efficiency is medium).

Classification is to derive the maximum a posteriori $P(C_i|X)$, then:

$$P(Under\ capacity|X) = P(X|Under\ capacity)\ x\ P(Under\ capacity)$$
$$= (0.24\ x\ 0.32\ x\ 0.40\ x\ 0.47\ x\ 0.50\ x\ 0.28\ x\ 0.81\ x\ 0.68)$$
$$x\ 0.41 = 0.00046$$
$$P(Sufficient\ capacity|X) = P(X|Sufficient\ capacity)\ x\ P(Sufficient\ capacity)$$
$$= (0.18\ x\ 0.22\ x\ 0.23\ x\ 0.48\ x\ 0.53\ x\ 0.21\ x\ 0.75\ x\ 0.76)$$
$$x\ 0.37 = 0.00010$$
$$P(Excess\ capacity|X) = P(X|Excess\ capacity)\ x\ P(Excess\ capacity)$$
$$= (0.05\ x\ 0.20\ x\ 0.04\ x\ 0.05\ x\ 0.05\ x\ 0.01\ x\ 0.36\ x\ 0.03)$$
$$x\ 0.21 = (1.08\ E - 10)\ x\ 0.21$$

Based on the a posteriori, if order X is assigned to the sewing line with the above characteristics, the probability of the order failing to meet the customer's production volume and delivery time requirements is highest, because the a posteriori has the highest score.

In practice, this result demonstrates how the model serves as a quantitative decision-support tool tailored to specific orders. During the order intake and production planning phase for a sewing line, production managers can input the new order's attributes and the designated line's technical specifications. The model then outputs posterior probabilities for three capacity states: under capacity, sufficient capacity, and excess capacity. Instead of providing a single point estimate for average output, the model yields a probability-based classification that reflects the interplay between order characteristics and sewing line specifications, thereby signaling potential risks before production actually begins. In addition, the model's rapid computation and reproducibility facilitate the testing of various production planning scenarios. By re-running order data across different line configurations, such as adjusting the number of workers, average skill levels, or line-balancing efficiency, the production manager can identify the setup that shifts the posterior probability toward a state of sufficient or excess capacity. Consequently, the production manager can not only decide whether to accept an order but also determine how to reallocate resources to meet output targets and delivery schedules.

The novelty of this approach becomes evident when compared with the experience-based forecasting methods that remain prevalent in many clothing enterprises. Traditional practices often rely on the subjective judgment of sewing line managers and historical average output data; such methods are inherently subjective, difficult to replicate, and heavily dependent on individual expertise. They typically treat orders as homogeneous, resulting in binary decisions (accept or reject) without quantifying the associated uncertainty. In practice, the model transforms accumulated experience into a data-driven operational tool, helping standardize and scale the application of expert knowledge while ensuring consistency in handling diverse, increasingly flexible orders.

5. Conclusion

This research developed and evaluated a data-driven model to forecast the production capacity of sewing lines in the clothing industry, employing a Naive Bayes classifier to link order characteristics with production line parameters. Instead of treating the 5M+1E framework merely as a qualitative tool for root cause analysis, the study restructured its elements into two groups, including order characteristics and production line characteristics, which are practically measurable during operations, using them as inputs for the forecasting model. This approach establishes a practical bridge between a widely recognized quality management diagnostic framework and quantitative data mining processes. In contrast to the company's existing method, which estimates capacity based on historical average output while overlooking the specific complexity of individual orders, the proposed model forecasts capacity at three levels: under capacity, sufficient capacity, and excess capacity for each order-production line pairing. The model generates probabilistic outputs, directly supporting decisions regarding order acceptance or rejection and resource reallocation amidst various uncertainties.

The model was trained and evaluated on data from 463 historical orders at a Vietnamese clothing company, using cross-validation. Results indicate an overall classification accuracy of 68.3%. Performance varied significantly across the groups: the “Excess Capacity” group was predicted with the highest reliability (true positive: 0.94; precision: 0.98), whereas the “Under Capacity” group (true positive: 0.65; precision: 0.64) and, in particular, the "Sufficient Capacity" group (true positive: 0.58; precision: 0.58) proved more difficult to classify accurately. One limitation of the study is that the model was developed based on a single dataset comprising 463 orders from a specific factory production line. The limited sample size, low product diversity, and scope restricted to a single factory affected the statistical reliability for groups with smaller sample sizes, as well as the model's applicability to other product types and factories. Future efforts to improve model accuracy and practical applicability could involve expanding the dataset in both size and product diversity to enhance reliability for capacity groups that yielded lower prediction performance. Additionally, the

Naive Bayes approach could be evaluated and compared against other machine learning algorithms, such as decision trees (J48), random forests, and ensemble learning methods. Furthermore, incorporating additional predictive variables, such as dynamic line balancing efficiency or worker learning effects, could more accurately reflect the specific characteristics of the sewing production line.

Acknowledgment

This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under a project within the framework of the Program titled "Strengthening the capacity for education and basic scientific research integrated with strategic technologies at VNU-HCM, aiming to achieve advanced standards comparable to regional and global levels during the 2025-2030 period, with a vision toward 2045", under grant number: CB2025-20-17.

References

[1] Liu, N., Chow, P. S., & Zhao, H. (2020). Challenges and critical successful factors for apparel mass customization operations: recent development and case study. *Annals of Operations Research*, *291*(1), 531–563.
[2] Fani, V., Antomarioni, S., Bandinelli, R., & Bevilacqua, M. (2023). Data-driven decision support tool for production planning: a framework combining association rules and simulation. *Computers in Industry*, *144*, 103800.
[3] Chowdhury, M. M. H., Rahman, S., Quaddus, M. A., & Shi, Y. (2023). Strategies to mitigate barriers to supply chain sustainability: an apparel manufacturing case study. *Journal of Business & Industrial Marketing*, *38*(4), 869–885.
[4] Palade, D., & Møller, C. (2023). Guiding digital transformation in SMEs. *Management and Production Engineering Review*, 14(1), 105–117.
[5] Yao, X., Almatooq, N., Askin, R. G., & Gruber, G. (2022). Capacity planning and production scheduling integration: improving operational efficiency via detailed modelling. *International Journal of Production Research*, *60*(24), 7239–7261.
[6] Wang, D., & Shao, X. (2024). Research on the impact of digital transformation on the production efficiency of manufacturing enterprises: Institution-based analysis of the threshold effect. *International Review of Economics & Finance*, *91*, 883–897.
[7] Joy, R. A., Hawlader, M. S., Rahman, M. S., Hossain, M. R., Shamim, S. I., & Mahmud, H. (2024). Improving quality, productivity, and cost aspects of a sewing line of apparel industry using TQM approach. *Mathematical problems in engineering*, *2024*(1).
[8] Banduka, N., & Aljinović Meštrović, A. (2025). Statistical Evaluation of Failure Mode Drivers via the 5M+ 1E Framework for Enhanced Reliability Control in Automotive Interior Manufacturing. *Journal of Engineering Management and Systems Engineering*, *4*(3), 176–187.
[9] Teshome, M. M., Meles, T. Y., & Yang, C. L. (2024). Productivity improvement through assembly line balancing by using simulation modeling in case of Abay garment industry Gondar. *Heliyon*, *10*(1).
[10] Park, H., & Kim, S. (2025). Development of a quantitative evaluation method for the skill level of sewing operators. *International Journal of Clothing Science and Technology*, *37*(5), 889–902.
[11] Begum, S., Akash, M. A. S., Khan, M. S., & Bhuiyan, M. R. (2024). A framework for lean manufacturing implementation in the textile industry: A research study. *International Journal of Science and Engineering*, *1*(04), 17–31.
[12] Amin, A., Bhuiyan, M. R. I., Hossain, R., Molla, C., Poli, T. A., & Milon, M. N. U. (2024). The adoption of Industry 4.0 technologies by using the technology organizational environment framework: The mediating role to manufacturing performance in a developing country. *Business Strategy & Development*, *7*(2).
[13] Powell, D. J., Laubengaier, D. A., Tortorella, G. L., Saabye, H., Antony, J., & Cagliano, R. (2024). Digitalization in lean manufacturing firms: a cumulative capability development perspective. *International Journal of Operations & Production Management*, *44*(6), 1249–1278.
[14] Yazdi, M. (2024). Maintenance strategies and optimization techniques. In *Advances in computational mathematics for industrial system reliability and maintainability* (pp. 43–58). Cham: Springer Nature Switzerland.
[15] Li, H., Kong, F., Kong, L., & Chen, T. (2023). Evaluating sewing operation complexity and its influence on sewing operation quality. *Heliyon*, *9*(3).
[16] Örüm Aydin, A., Zajonz, F., Günther, T., Dermenci, K. B., Berecibar, M., & Urrutia, L. (2023). Lithium-ion battery manufacturing: Industrial view on processing challenges, possible solutions and recent advances. *Batteries*, *9*(11), 555.

[17] Bizuneh, B., & Mamecha, R. (2025). Evaluation of an Industrial Sewing Workstation Design and Its Impact on Perceived Productivity. *Journal of Engineering*, *2025*(1).

[18] Gong, W. (2024). Optimization Study of the Production Layout in the Trouser Workshop of GR Company. *Journal of Business and Marketing*, *1*(3), 356–360.

[19] Jayarathna, S. M. D. Y. (2014). An empirical investigation of the impact of employee motivation and productivity of executive employees with special reference to the apparel industry in Sri Lanka. *International Journal of Science and Research*, *3*(3), 749-754.

[20] Lingkon, M. L. R., Saha, P. K., Manzid, A. A., Hasan, M. N., & Mahalanobish, S. K. (2024). Reducing sewing defects to increase productivity in the apparel industry of Bangladesh by integrating lean methodology. *International journal of research in industrial engineering*, *13*(2), 166–187.

[21] Paresashvili, N., Okruashvili, N., Tikishvili, M., Chitaladze, K., & Avsajanishvili, E. (2024). Influence of factors of a healthy working environment on employee labor productivity (on the example of higher educational institutions of Georgia).

[22] Somanathan, E., Somanathan, R., Sudarshan, A., & Tewari, M. (2021). The impact of temperature on productivity and labor supply: Evidence from Indian manufacturing. *Journal of Political Economy*, *129*(6), 1797-1827.

[23] Tyagi, S. K., & Sharma, B. K. (2011). Data mining tools and techniques to manage the textile quality control data for strategic decision making. *International Journal of Computer Applications*, 13(4), 26–29.

[24] Yildirim, P., Birant, D., & Alpyildiz, T. (2017). Data mining and machine learning in textile industry. *WIREs Data Mining and Knowledge Discovery*, 8(1).

[25] Gowtham, K. R., Shetty, M., & Shetty, S. B. (2025, February). A Forecast of Consumer Behaviour using the Naivesbayes Algorithm. In *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)* (pp. 1840-1846). IEEE.

[26] Lowd, D., & Domingos, P. (2005, August). Naive Bayes models for probability estimation. In *Proceedings of the 22nd international conference on Machine learning* (pp. 529-536).

[27] Elmaghraby, S. E. (2010). Production capacity: Its bases, functions and measurement. In *Planning Production and Inventories in the Extended Enterprise: A State of the Art Handbook, Volume 1* (pp. 119–166). New York, NY: Springer US.

[28] Ramirez-Gutierrez, H. L., & Gómez-Marin, C. G. (2024). Methodology for Diagnosing the Impact of Operations Management Elements in Technical and Vocational Education and Training in the Garment Industry. *Management and Production Engineering Review*.