International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Forecasting New Product Success: A Methodology Combining Consumer Preference Analysis and Machine Learning Prediction Models

Doohee Chung*[1], Chan Gyu Lee[2], Sungmin Yang[3]

Submitted:10/08/2024 **Accepted:**20/09/2024 **Published:**29/09/2024

Abstract*:* In an increasingly competitive market with shortening product lifecycles, developing successful new products is more critical than ever. Traditional methods, which assess product performance post-launch, often lead to missed opportunities and inefficiencies. This study introduces a novel methodology that combines consumer preference analysis with advanced predictive modeling using Gaussian process regression to forecast new product success. By integrating the Product Differentiation Index with the Demand Creation Index and incorporating user satisfaction data from the KANO model, this approach offers a robust tool for predicting market demand before a product even hits the shelves. Tested in the dynamic smartwatch industry, the model demonstrated high accuracy, with a MAPE value of 0.13, and identified pulse detection as the feature most likely to drive sales in future products. This innovative methodology not only predicts early-stage demand but also equips companies with the insights needed to make strategic, data-driven decisions that maximize market impact and profitability.

Keywords: *Predictive new product development, Demand Forecast, Gaussian Process Regression, KANO, Product Differentiation*

1. Introduction

New product development is essential for the sustainable growth of companies, as it directly influences sales, profitability, and competitiveness [17]. In a rapidly evolving market with shortening product lifecycles, the importance of an effective new product development strategy is paramount [24]. Numerous studies have explored methods to enhance the effectiveness of new product development, focusing on processes from idea conception to commercialization [3], stage-gate evaluations, opportunity discovery [20], and integration with organizational management [5].

However, the complexity and uncertainty of markets make it challenging to accurately gauge consumer needs, limiting companies' ability to predict the impact and costs associated with new product development. Traditionally, many product planning decisions have relied on the intuition of developers, which can lead to a mismatch with market needs [29]. Most existing methodologies are reactive, assessing sales performance only after product launch, making it difficult to recover from failures due to the significant time and resources invested. Therefore, this study posits that a proactive approach—predicting sales performance prior to production—would be more effective in enhancing the success rate of new product development.

In recent years, there has been a shift toward using data science approaches in new product development to improve prediction accuracy and manage complexity [9, 7]. Data science can uncover patterns within complex data, enabling more precise future predictions [25]. When applied to product development, these approaches can facilitate strategic product launches with higher success probabilities by enabling predictive planning.

While previous studies, such as those by Van Steenbergen & Mes [41] and Afrin & Monplaisir [1], have developed models for demand forecasting, they often fail to detail the product features that maximize demand creation. Existing methodologies may also struggle to incorporate new features that have not yet been introduced to the market.

This study addresses these limitations by calculating a Product Differentiation Index (PDI) based on consumer satisfaction with both existing and new features. A machine learning model is then used to capture the relationship between this index and the Demand Creation Index, which measures changes in sales of new products compared to existing ones. This model enables early prediction of the demand creation potential of new features and identifies the product profiles that maximize future demand, as demonstrated in a case study on smartwatch development.

1 Corresponding Author, School of Global Entrepreneurship and Information Communication Technology (ICT), Handong Global University – 05319, Republic of Korea
Email: profchung@handong.edu
ORCID ID : 0000-0002-7413-0778
2 Graduate School of Data Science, Seoul National University –05319, Republic of Korea
3 Technology Management Economics and Policy Program, College of Engineering, Seoul National University – 05319, Republic of Korea

2. Research Methodology

New products typically incorporate added or improved features compared to existing ones. These differential features lead to variations in demand, such as sales performance. Therefore, the degree of demand creation can be predicted using a differentiation index that compares new products to existing ones, a concept supported by previous studies [1]. This research proposes a method for predicting demand by modeling the relationship between a differentiation index, which reflects the degree of feature differentiation, and an index that reflects changes in demand. Specifically, this method incorporates market response through the KANO model in calculating the differentiation index.

Figure 1 illustrates the framework of the predictive new product development methodology proposed in this study. To calculate the differentiation index, customer satisfaction coefficients derived from the KANO model and the degree of differentiation for each attribute are used as weights. The demand creation prediction model is then built by fitting the relationship between the Product Differentiation Index and the Demand Creation Index. The model uses Gaussian process regression, a method effective for analyzing unstructured and non-parametric data, especially suitable for modeling with limited data as in this study. This model predicts initial sales volumes for various new product scenarios and identifies the product profile with the highest potential sales performance. The detailed methodology for each step is as follows.

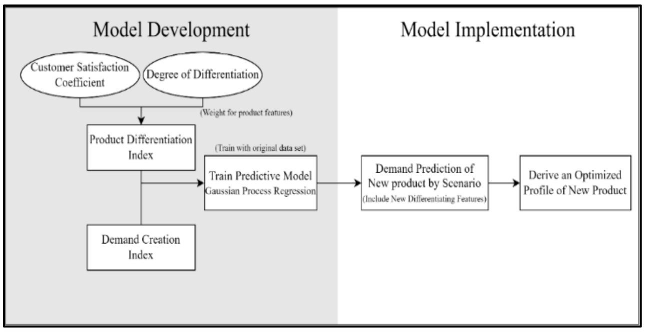

Fig. 1. Framework of Predictive Methodology for New Product Development

2.1. Kano

An initial step in building a predictive new product development model is to calculate the differentiation index, which measures how differentiated the new product's features are compared to its predecessor. This process involves subdividing the product's features, determining the degree of differentiation for each feature, and assigning weights to reflect their importance. Unlike the expert evaluations used in previous studies like Afrin & Monplaisir [1], which may not align with market perceptions, this study uses the KANO model to capture consumer responses more accurately and reflect the importance of each feature from the consumer's perspective.

The KANO model, introduced by Kano [19], systematically explains consumer reactions by distinguishing between satisfaction and dissatisfaction. This model is widely used in new product development to analyze consumer responses to various product features [36, 11, 2]. It evaluates each feature of a product based on five quality factors derived from survey data collected from users.

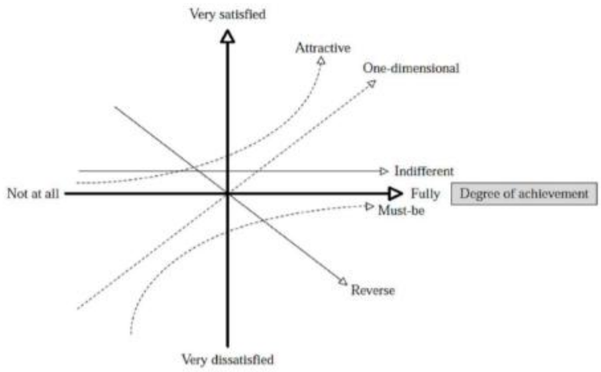

Fig. 2. KANO Model

Table 1. KANO Quality Element

List of Quality Element	*Explanation*
Attractive Quality Element (A)	Features of the product that give customer unexpected satisfaction. If satisfied, customer satisfaction increases significantly, and if not customer satisfaction decreases little.
One-Dimensional Quality Element (O)	The most common type of features of the product. If satisfied, customer satisfaction increases, and if not customer satisfaction decreases.
Must-be Quality Element (M)	Features of a product that customer take for granted to be satisfied. If not satisfied, customer satisfaction decreases significantly, and if satisfied, customer satisfaction increases little.
Indifferent Quality Element (I)	Features of the product that customer is not interested in. Whether satisfied or not, they do not change customer satisfaction.
Reverse Quality Element (R)	Features of the product in the opposite concept of a One-dimensional Quality Element. If satisfied, customer satisfaction decreases, and if not customer satisfaction increases.
Skeptical Quality Element (S)	Features of the product that the customer is suspected of understanding. Whether satisfied or not, customer is satisfied or dissatisfied in all situations, Not used to calculate satisfaction coefficients.

The questionnaire for KANO analysis is conducted in such a way that positive and negative questions are asked simultaneously for each function or element of the product. A positive question (Functional Question) is a question that asks how consumers feel when a product satisfies a corresponding property or function, and a negative question (Dysfunctional Question) is a question asking how a consumer feels when a product does not satisfy a corresponding property or function. The response options of KANO questionnaire consist of ‘Like (1)’, ‘Expect (2)’, ‘Neutral (3)’, ‘Tolerate (4)’, and ‘Dislike (5)’. The survey respondent selects one of the responses to each of the positive and negative questions, which fall into one of 25 categories listed in Table 2.

Table 2. KANO Quality Element Table

Quality Element		Dysfunctional Question				
		1	2	3	4	5
Functional Question	1	S	A	A	A	O
	2	R	I	I	I	M
	3	R	I	I	I	M
	4	R	I	I	I	M
	5	R	R	R	R	S

Q: Skeptical Quality Element, A: Attractive Quality Element,
I: Indifferent Quality Element, M: Must-Be Quality Element,
O: One-Dimensional Quality Element, R: Reverse Quality Element

After classifying the quality element of each product, the KANO model applies customer satisfaction coefficient formula developed by Timko [40] to numericalize customer satisfaction for each product attribute.

$$Satisfaction\ Coefficient = \frac{A+O}{A+O+M+I} \quad (1)$$

This is derived by calculating the proportion of attractive, one-dimensional quality elements in the sum of all quality elements responses. In this way, it is possible to grasp the satisfaction coefficient for each attribute of the product. The satisfaction coefficient for each attribute is used as a weight for calculating the differentiation index of a new product.

2.2. Product Differentiation Index

After determining the weights for each product attribute using the KANO model, the next step is to calculate the degree of differentiation for each attribute. Product differentiation typically involves enhancing existing features or adding new ones [21], and in IT products, innovative features can significantly impact the market [28]. To accurately calculate the differentiation index, it is essential to assess the degree of differentiation between the new product and its predecessor for each feature.

This study categorizes features as 'No-Differentiation' if they are unchanged or downgraded from the predecessor, 'Weak Differentiation' if improved by 20%, and 'Strong Differentiation' if improved by more than 20% or if a completely new feature is added. When features cannot be measured numerically, expert evaluation is used. The weights of 0, 0.2, and 0.7 are assigned to 'No-Differentiation,' 'Weak Differentiation,' and 'Strong Differentiation,' respectively, following existing research methods. To calculate the new product's differentiation index, the customer satisfaction coefficient for each feature and the degree of differentiation are combined using the weights discussed above.

Table 3. Differentiation Index by Feature (Sample)

Feature	*Predecessor Product*	*New Product*	*Satisfaction Coefficient*	*Degree of Differentiation*	*Differentiation Index by Feature*
A	Standard	15% upgrade	0.12	0.2	0.024
B	None	Newly added	0.32	0.7	0.224
C	Standard	70% upgrade	0.14	0.7	0.098
D	Standard	10% upgrade	0.24	0.2	0.048
Total Differentiation Index of New Product					0.250

For example, in the case of smartwatches, a new product might include features identical to those in the existing series, some slightly or significantly upgraded features, and entirely new features. To calculate the differentiation index of the new product, the customer satisfaction coefficient and the degree of differentiation for each feature are determined. These two values are then multiplied to obtain the differentiation index (C_i) for each feature. The overall Product Differentiation Index (PDI) is then calculated by summing the differentiation indexes of all features, providing a comprehensive numerical value that represents the difference between the new product and its predecessor.

$$C_i = Satisfaction\ Coefficient_{\ i} \times Degree\ of\ Differentiation_i \quad (2)$$

$$i = 1, \dots, n$$

In the above formula, n is the number of differentiated features in the new product among those selected as the core features of the product.

$$Product\ Differentiation\ Index = \sqrt{\sum_{i=1}^{n} (C_i)^2} \quad (3)$$

The next step is to normalize the PDI value to construct a model with the Demand Creation Index. The total differentiation index is calculated using the Euclidean distance between the features of the predecessor and the new product. If these values vary widely, it could negatively impact model performance [6]. To normalize the differentiation index, maxPDI is calculated by assuming that all key features of the new product have maximum improvement compared to the predecessor. Conversely, minPDI is calculated assuming minimal or no improvement in core features, resulting in a value of 0. The normalized PDI (z value) is then derived using PDI, maxPDI, and minPDI, which serves as an input for the machine learning model.

$$z = \frac{PDI - minPDI}{maxPDI - minPDI} \quad (4)$$

2.3. Demand Creation Index

To predict demand using the previously derived PDI of new products, the study introduces the Demand Creation Index (DCI), which measures the difference in demand between a new product and its predecessor. DCI is calculated as the ratio of the initial sales volume of the new product to that of the predecessor during the same sales period. For IT devices, where product lifecycles are short, sales within the first year often represent the majority of the product's performance, so early sales data, such as from the first month or quarter, is typically used for evaluation [33].

The DCI formula indicates that a DCI of 0 means the sales volume is unchanged from the predecessor, a DCI greater than 0 means the new product outperforms the predecessor, and a DCI less than 0 indicates lower sales. By modeling the relationship between DCI and PDI, the sales volume of a new product can be predicted using the PDI of its feature combination.

$$Demand\ Creation\ Index = \frac{New\ Product\ Demand - Predecessor\ Product\ Demand}{Predecessor\ Product\ Demand} \quad (5)$$

2.4. Gaussian Process Regression

To predict the DCI of a new product using the previously derived PDI, a suitable algorithm is needed to accurately model the relationship between the two. While some studies suggest that large product differences positively impact demand [37, 23, 8], others argue that smaller changes are more beneficial [32, 27]. Desai et al. [10] proposed that both too large and too small differences can negatively affect demand, with sales peaking at an optimal level of differentiation.

These conflicting findings highlight that the relationship between product differentiation and demand is nonlinear, particularly in the complex and irregular real market data. Therefore, a linear model would be insufficient. In selecting a model to explain the relationship between product differentiation and demand creation, three criteria are considered: the model must capture nonlinear relationships, it should be non-parametric to accommodate irregular data distributions [30, 39], and it must perform well with limited data due to the scarcity of historical new product data.

Gaussian Process Regression (GPR) is chosen as the optimal model, meeting these criteria. GPR is a non-parametric model that handles nonlinear problems effectively and performs well even with small datasets [44]. Unlike traditional machine learning models, GPR relies on a small number of parameters, generating predictions by introducing an explicit basis function to model the latent variable.

$$f(X) = GP(m(X), k(X, X')) \tag{6}$$

In the Gaussian Process Regression (GPR) formula, m(X) represents the mean, and k(X,X′) represents the covariance. GPR uses these mean and covariance functions to determine the distribution that corresponds to the confidence interval of the predicted value. It predicts outcomes by deriving the variance of this distribution. The covariance function serves as a kernel parameter to measure similarity among all training data points. The kernel function creates a posterior predictive distribution, giving more weight to predictions as similarity increases, thus deriving a value suited to the distribution. This approach allows GPR to perform better than general predictive models, especially when dealing with small datasets or complex, nonlinear market data [14]. Therefore, GPR is particularly effective in situations with limited data or when linear characteristics are difficult to determine.

$$y = hX'\beta + f(X) \tag{7}$$

Using this model, the x-axis is set as the PDI value and the y-axis is set as the DCI value to complete a predictive model that explains the effect of product differentiation on change in demand. Finally, by using Bayes' theorem in the corresponding equation, the optimal y value reflecting the existing results, the final predicted value of DCI is derived. The predicted DCI value derived in this way is multiplied by and added to the demand of the predecessor product to calculate the final result of the model. This becomes the first quarter predicted demand after the launch of the new product.

$$\begin{gathered} Initial\ Demand\ Prediction = \\ Predicted\ DCI\ \times Predecessor\ Product\ Demand + \\ Predecessor\ Product\ Demand \end{gathered} \tag{8}$$

2.5. Optimized New Product Profiles

Using the demand forecasting model built with the PDI and DCI through Gaussian process regression, it becomes possible to predict the sales volume of new products based on customized feature combinations. This approach helps identify the features that can maximize sales performance for the next new product.

The process to derive the optimal new product profile includes several steps. First, select the most recent existing product or the product targeted for innovation. Various scenarios are then created by configuring features to be added or improved in the new product compared to the selected product. The satisfaction coefficient for each new feature should be obtained in advance using the KANO model to apply the appropriate weight during modeling.

Next, the initial sales volume is predicted for each feature scenario using the prediction model. By comparing these sales volumes, the feature combination that maximizes sales is identified as a candidate for the new product profile. The product development manager can then determine the development direction and make decisions based on this optimized profile. This method is valuable because it allows for early identification of a product profile that can maximize sales, thereby improving the new product development process.

3. Case Study: Smartwatch Product Development

To verify the proposed methodology, this study applies it to smartwatch product development. Smartwatches, such as those from Apple, Samsung, Huawei, and Fitbit, are ideal for this method due to their regular release cycles and frequent feature innovations.

3.1. Data

In this study, data from Apple Watch and Samsung Electronics' Galaxy Watch are selected as the subject of analysis as they relatively have a more sufficient set of comparable data. The data includes Apple Watch Series 1~7 and Apple Watch SE, released from September 2014; Galaxy Watch Series 1~3, Galaxy Watch Active 1-2, and Galaxy Watch Classic launched from August 2018. The analysis was conducted based on the data from the Korean region. Many studies use data from the Korean market as Korea has a high level of maturity in IT technology and is used as a test bed for new IT products [22].

The data required for analysis include quarterly shipment data, data on features and specifications of smartwatches, and user satisfaction coefficient data for each feature. Since collecting quarterly shipment data for the Korean market was difficult, the data was obtained according to the following method. First, through Euromonitor [13], SA [38], and Gartner [15], information on the quarterly market size (total sales) of smartwatches in Korea is acquired. Afterward, Danawa Research [35], the company of product information and price comparison service, provided the quarterly sales ratio for each product. Based on this information, specific quarterly sales volume for each product could be inferred. Product feature data was obtained through each product's website and Versus [42]. Customer satisfaction coefficient data by product feature was obtained through a KANO-based survey. The data obtained through the Kano Survey consists of response data from a sample of 160 people in their teens to 60s. The distribution of respondents was 1.8% for 10s, 61.2% for 20s, 20% for 30s, 3.1% for 40s, and 13.75% for 50s and older. By gender, 56.8% were male and 43.2% were female. Based on these data, the PDI and DCI of a new product are calculated to build a predictive model. The specific method is as follows.

3.2. Product Differentiation Index

The Product Differentiation Index (PDI) measures how much a new product differs from its predecessor, which is essential for predicting its mfarket success. First, customer satisfaction coefficients are determined for each of the 20 selected smartwatch features using a KANO survey. Participants rated their satisfaction on a 5-point Likert scale, and these ratings were used to calculate the satisfaction coefficients for each feature.

Table 4. Classification of Quality Elements and Satisfaction Coefficient result of Features

Features	*Q*	*A*	*I*	*M*	*O*	*R*	*Satisfaction Coefficient*
Launching Price (Won)	2	71	39	10	31	7	0.6755
Weight (g)	2	40	16	31	70	1	0.7006
Battery (mAh)	3	45	12	9	89	2	0.8645
Display Resolution	1	32	32	18	76	1	0.6835
RAM	1	54	42	6	57	0	0.6981
Memory	0	55	50	10	44	1	0.6226
Design (Score)	1	25	8	7	119	0	0.9057
OS	0	51	37	9	63	0	0.7125
Wi-Fi	1	42	19	17	81	0	0.7736
Bluetooth	0	27	15	22	93	3	0.7643
Cellular Connectivity	3	14	5	58	80	0	0.5987
Electrical Heart Rate	0	58	21	12	68	1	0.7925
VO2max	1	70	28	8	51	2	0.7707
ECG app	2	72	40	6	37	3	0.7032
Stress Measurement	1	94	27	2	32	4	0.8129
Fall Detection	0	95	17	4	41	3	0.8662
Water Intake	3	61	13	23	60	0	0.7707
Calorie Intake	2	86	24	3	44	1	0.8280
Irregular Heart Rate Warnings	2	88	23	0	46	1	0.8535

Q: Skeptical Quality Element, A: Attractive Quality Element,
I: Indifferent Quality Element, M: Must-Be Quality Element,
O: One-Dimensional Quality Element, R: Reserve Quality Element

Next, the degree of differentiation for each feature compared to the predecessor is assessed. This differentiation is categorized into three levels: no differentiation (0), weak differentiation (0.2), and strong differentiation (0.7). For instance, if the operating system (OS) of the Apple Watch Series 4 shows significant improvement over Series 3, it would be assigned a strong differentiation weight of 0.7. In contrast, a slight design change might be classified as weak differentiation with a weight of 0.2.

The PDI for each feature is then calculated by multiplying the satisfaction coefficient by the degree of differentiation. The total PDI for the new product is obtained by summing these individual PDIs, providing a comprehensive measure of the product's overall differentiation compared to its predecessor. To ensure that the PDI is comparable across different products, it is normalized using maxPDI, which assumes strong differentiation across all features, and minPDI, which assumes no differentiation. This normalization process results in a standardized value (z) that is used in predictive modeling.

Table 5. Differentiation Index of Apple Watch 4 Vs. Apple Watch 3

Feature	*Satisfaction Coefficient*	*Degree of Differentiation*	*Differentiation Index by Feature*
Launching Price (Won)	0.6755	0	0
Weight (g)	0.7006	0.7	0.4904
Battery (Duration)	0.8645	0	0
Display resolution	0.6835	0.7	0.4785
RAM	0.6981	0.2	0.1396
Memory	0.6226	0.2	0.1245
Design (Score)	0.9057	0.2	0.1811
OS	0.7125	0.7	0.4988
Wi-Fi	0.7736	0	0
Bluetooth	0.7643	0	0
Cellular Connectivity	0.5987	0	0
Electrical Heart Rate	0.7925	0.7	0.5395
VO2max	0.7707	0	0
ECG app	0.7032	0	0
Stress Measurement	0.8129	0	0
Fall Detection	0.8662	0.7	0.6064
Multi-sports Mode	0.7707	0	0
Water Intake	0.8280	0	0
Calorie Intake	0.8535	0	0
Irregular Heart Rate Warnings	0.8671	0	0
Total Differentiation Index of New Product			1.2020

※ In Degree of Differentiation, Strong Differentiation is 0.7, Weak Differentiation is 0.2, and No-Differentiation is 0.

Table 6. Nomalization of Apple Watch 4 Differentiation Index

PDI	*maxPDI*	*minPDI*	*z*
1.2020	2.7492	0	0.4372

3.3. Demand Creation Index

To assess the rate of change in sales between new and predecessor smartwatches, it is first necessary to pair the initial demand data of both products. Given the short lifecycle of smartwatches, this study uses sales volume from the first quarter (four months) as the initial demand data for the forecasting model. The Demand Creation Index (DCI) is calculated by dividing the difference in initial demand between the new product and its predecessor by the initial demand of the predecessor. This DCI value serves as the dependent variable in the predictive model. For example, Table 7 demonstrates the calculation of the DCI for Apple Watch models, where the initial demand data of new and predecessor products are used to determine the percentage change in sales volume.

Table 7. Apple Watch Demand Creation Index Calculation

Model		*Predecessor Product Demand*	*New Product Demand*	*DCI*
New product	*Predecessor product*			
Apple watch 4	Apple watch 3	1092030	873727	-0.20
Apple watch 5	Apple watch 3	1092030	3156387	1.89
Apple watch 6	Apple watch 3	1092030	2746727	1.515
Apple watch SE	Apple watch 3	1092030	1610277	0.475
Apple watch 7	Apple watch 3	1092030	3786952	2.468
Apple watch 5	Apple watch 4	873727	3156387	2.613
Apple watch 6	Apple watch 4	873727	2746727	2.144
Apple watch SE	Apple watch 4	873727	1610277	0.843
Apple watch 7	Apple watch 4	873727	3786952	3.334
Apple watch 6	Apple watch 5	3156387	2746727	-0.13
Apple watch SE	Apple watch 5	3156387	1610277	-0.49

Apple watch 7	Apple watch 5	3156387	3786952	0.20
Apple watch SE	Apple watch 6	2746727	1610276	-0.414
Apple watch 7	Apple watch 6	2746727	3786952	0.379
Apple watch 7	Apple watch SE	1610276	3786952	1.352

Once the PDI and DCI for a new smartwatch are calculated, the next step is to develop a predictive model that fits the relationship between these two indices. To ensure robust predictive modeling, 23 complete data points were selected after excluding outliers, with 80% of the data used for training and 20% for testing. The test data is carefully selected to ensure diverse smartwatch predictions, avoiding repetition of the same product.

Gaussian process regression is then applied to build the predictive model. This method predicts the DCI by introducing the latent variable of the Gaussian process along with an explicit basis function for the PDI. The study utilizes the gp function from the GauPro package in R 4.0.2, setting up a Gaussian process regression model framework. An exponential kernel function is used to measure similarity across all training data points. This process enables the Gaussian process regression to generate a prediction trend curve, which is then used to predict the DCI for the test data, ensuring accurate and high-quality predictions.

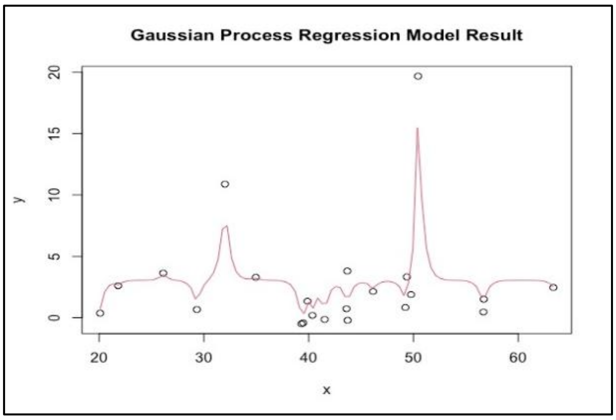

Fig. 3. Gaussian Process Regression Model Curve

3.4. Evaluation

This study uses Gaussian process regression to predict DCI and utilizes the MSE (Mean Squared Error) and MAPE (Mean Absolute Percentage Error) indicators to measure the accuracy of the model. MSE means the average of the squared difference between each predicted value and the actual value derived by the model. When this index is closer to zero, the prediction error is lower and the accuracy is higher.

$$MSE = \frac{1}{n}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (9)$$

The MAPE indicator is used to measure the accuracy of the predicted initial demand. MAPE is an index that divides the prediction error by the actual value and expresses it as a percentage value according to the data size. As with MSE, when the number is low, it means a more accurate model with lower prediction error.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{N}\left|\frac{A_t - F_t}{A_t}\right| \quad (10)$$

4. Result

4.1. Prediction Accuracy

The results of the prediction model using the case of the smartwatch are as follows.

Table 8. Initial Demand Prediction Result

Testing Model		*DCI*	*DCI*	*Initial Demand*	*Initial Demand*
New Product	*Predecessor Product*	*(Actual)*	*(Predicted)*	*(Actual)*	*(Predicted)*
Apple Watch 5	Apple Watch 4	2.61	3.01	3,156,387	3,503,821
Apple Watch 6	Apple Watch 3	1.51	1.12	2,746,727	2,320,689
Apple Watch 7	Apple Watch SE	1.35	1.20	3,786,952	3,553,003
Galaxy Watch 4	Galaxy Watch 3	3.29	3.05	11,111,038	10,497,587

The DCI prediction result using the Gaussian process regression method recorded an accuracy of 0.097 based on the MSE indicator, and that the overall mean squared error showed high accuracy. The closer the predicted value of DCI is to the actual value, the more directly it affects the initial sales prediction for new products. Therefore, the final result of the prediction model, that is, the initial sales volume prediction performance also recorded 9.6% based on the MAPE, confirming relatively high prediction performance.

As this study uses small quantity of data in modeling process, it is necessary to measure performance more clearly. To this end, the robustness of the model results was confirmed through cross-validation technique. The test was repeated with an additional randomized subset of the same size as the base test from the entire dataset. As a result of 5-fold cross-validation (5-folds CV), in Table 9, the average performance of all tests was 0.205 MSE and 12.2% MAPE, maintaining a relatively high level of accuracy.

Table 9. 5-fold Cross Validation of Predictive Model Performance

Cross Validation	*MSE*	*MAPE (%)*
CV 1	0.097	9.568
CV 2	0.240	12.183
CV 3	0.312	15.795
CV 4	0.191	12.574
CV 5	0.187	11.359

4.2. Optimized New Product Profiles

Using the predictive model, sales volumes for new product scenarios with added features were forecasted. Apple Watch 7 served as the predecessor product, and experts identified three new features: simultaneous translation, pulse detection, and virus measurement. These were incorporated into three different product profiles, with satisfaction coefficients obtained from the KANO survey.

Table 10 shows predicted first-quarter sales for each profile and the sales impact of each feature. The impact was calculated by comparing sales predictions with and without the feature using the Gaussian process regression model.

The pulse detection feature was predicted to generate the highest initial demand. It uses the watch to detect the user's pulse, offering health insights through an oriental medicine approach, potentially leading to a concept called the "Oriental Watch." This feature is likely to be well-received in the market. Simultaneous translation and virus measurement were also predicted to boost sales, though to a lesser degree.

Table 10. Demand Prediction & Demand Contribution Result of New Product Profiles

Scenario	*New Product Profiles*	*Predicted Demand (Initial Quarter)*	*Demand Contribution of New Feature*
1	Apple Watch 7 + Simultaneous Translation	5,533,184	0.461
2	Apple Watch 7 + Pulse Detection	6,224,969	0.644
3	Apple Watch 7 + Virus Measurement	5,484,724	0.448

5. Conclusion

Companies often struggle to develop effective new product strategies due to the rapidly changing market and reliance on subjective expert opinions, making it difficult to create a new product profile that maximizes sales [43]. This study addresses this challenge by developing a model that predicts how differences in product features affect initial sales volume, leading to a new methodology for developing products that maximize demand.

A predictive model was constructed using Gaussian process regression to model the relationship between the Product Difference Index (PDI) and the Demand Creation Index (DCI). This model was validated using data from the smartwatch market, specifically from Apple and Samsung. The model's accuracy was confirmed to be relatively high. By adding new features to the Apple Watch 7, the study tested various product profiles. The pulse detection feature was found to contribute most significantly to initial sales, followed by simultaneous translation and virus measurement features. This approach allows companies to proactively identify product profiles that enhance demand and improve the development process.

The study is significant for several reasons. First, it introduces an objective, data-driven approach to new product development, moving away from methods based heavily on subjective experience. Second, by incorporating market response through the KANO model as a weighting factor, the model better reflects user satisfaction and market needs, distinguishing it from other data science approaches. Finally, this methodology allows companies to predict sales early in the development process, transitioning from reactive to proactive product design aimed at maximizing sales.

However, the study has some limitations. The smartwatch market is relatively new, limiting the amount of available data. As the market matures, more sophisticated models can be developed with higher-quality data. Additionally, sales data for some products had to be estimated, which may affect accuracy. To address these limitations, future research plans to incorporate data from newer models, such as the Apple Watch 8, 9, and 10, to update and refine the predictive model. This will enhance the model's relevance and applicability to current market conditions. Furthermore, future models could improve by incorporating more detailed data and considering additional variables such as advertising expenditures and macroeconomic indicators, which would enhance predictive accuracy. Acknowledging these limitations while outlining steps for future research will ensure that the study remains robust and relevant in the context of ongoing technological advancements.

Acknowledgements

The authors would like to express their gratitude to **Impactive AI** for their generous funding support, which made this research possible. Their contribution has been instrumental in the development and completion of this study.

Author contributions

Doohee Chung: Conceptualization, Project Administration, Resources, Software, Supervision, Validation, Writing & Editing
Chan Gyu Lee: Software, Methodology, Investigation, Visualization, Writing & Editing
Sungmin Yang: Software, Methodology, Investigation, Visualization, Writing & Editing

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] K. Afrin, B. Nepal and L. Monplaisir, “A data-driven framework to new product demand prediction: Integrating product differentiation and transfer learning approach,” Expert Systems with Applications, vol. 108, pp. 246-257, 2018.

[2] R. Aliyu, A. M. T. Arifin, R. H. A. Haq, M. F. Hassan, M. N. A. Rahman, A. E. Ismail, M. Z. Rahim, M. R. Ibrahim, M. A. Azlan and M. S. Ahmad, “An integration of kano model and quality function deployment technique-a case study using sport earphone,” In Journal of Physics: Conference Series, vol. 1150, no. 1, pp. 012025, 2019.

[3] A. Booz, “Hamilton. (1982). New products management for the 1980s,” Booz, Allen & Hamilton, 1982.

[4] M. Belyaev, E. Burnaev and Y. Kapushev, “Exact inference for gaussian process regression in case of big data with the cartesian product structure,” arXiv preprint arXiv:1403.6573, 2014.

[5] H. J. Boeddrich, “Ideas in the workplace: a new approach towards organizing the fuzzy front end of the innovation process,” Creativity and innovation management, vol. 13, no. 4, pp. 274-285, 2004.

[6] D. Borkin, A. Némethová, G. Michaľčonok and K. Maiorov, “Impact of data normalization on classification model accuracy,” Research Papers Faculty of Materials Science and Technology Slovak University of Technology, vol. 27, no. 45, pp. 79-84, 2019.

[7] K. Y. Chan, C. K. Kwong, P. Wongthongtham, H. Jiang, C. K. Y. Fung, B. Abu-Salih, Z. Liu, T. C. Wong and P. Jain, “Affective design using machine learning: a survey and its prospect of conjoining big data,” International Journal of Computer Integrated Manufacturing, vol. 33, no. 7, pp. 645-669, 2020.

[8] Y. L. Cheng, “Vertical product differentiation under demand uncertainty,” Economic Modelling, vol. 36, pp. 51-57, 2014.

[9] K. Christensen, S. Nørskov, L. Frederiksen and J. Scholderer, “In search of new product ideas: Identifying ideas in online communities by machine learning and text mining”, Creativity and Innovation Management, vol. 26, no. 1, pp. 17-30, 2017.

[10] P. Desai, S. Kekre, S. Radhakrishnan and K. Srinivasan, “Product differentiation and commonality in design: Balancing revenue and cost drivers,” Management Science, vol. 47, no. 1, pp. 37-51, 2001.

[11] G. Dominici, V. Roblek, T. Abbate and M. Tani, “Click and drive: Consumer attitude to product development: Towards future transformations of the driving experience,” Business process management journal, vol. 22, no. 2, pp. 420-434, 2016.

[12] Y. Esmer and O. Saylan, "ROLE OF CUSTOMER INVOLVEMENT IN NEW PRODUCT DEVELOPMENT PROCESS OF INDUSTRIAL ENTERPRISES: A CONCEPTIONAL RESEARCH," Verimlilik Dergisi, no. 2, pp. 179-193, 2020.

[13] *Euromonitor*, Wearable Electronics in South Korea, 2021, Accessed on: November 1, 2022. https://www.euromonitor.com/wearable-electronics-in-south-korea/report

[14] W. Fang, S. Zhang, H. Huang, S. Dang, Z. Huang, W. Li, Z. Wang, T. Sun and H. Li, "Learn to make decision with small data for autonomous driving: deep gaussian process and feedback control," *Journal of Advanced Transportation*, vol. 2020, no. 1, pp. 8495264, 2020.

[15] *Gartner*, "Gartner: Fueling the Future of Business," 2018. [Online]. Available: https://www.gartner.com/en

[16] *Gartner*, "South Korea Talent Intelligence Library," 2021. [Online]. Available: https://www.gartner.com/en/documents/4002351

[17] S. I. Hallstedt, O. Isaksson and A. Öhrwall Rönnbäck, "The need for new product development capabilities from digitalization, sustainability, and servitization trends," *Sustainability,* vol. 12, no. 23, pp. 10222, 2020.

[18] T. E. Johnsen, "Supplier involvement in new product development and innovation: Taking stock and looking to the future," *Journal of Purchasing and Supply Management,* vol. 15, no. 3, pp. 187-197, 2009.

[19] N. Kano, N. Seraku, F. Takahashi and S. Tsuji, "Attractive quality and must-be quality," *Journal of the Japanese Society for Quality Control*, vol. 14, pp. 39-48, 1984.

[20] P. Koen, G. Ajamian, R. Burkart, A. Clamen, J. Davidson, R. D'Amore, C. Elkins, K. Herald, M. Incorvia, A. Johnson, R. Karol, R. Seibert, A. Slavejkov and K. Wagner, "Providing clarity and a common language to the "fuzzy front end"," *Research-technology management,* vol. 44, no. 2, pp. 46-55, 2001.

[21] H. L. Lee and C. S. Tang, "Modelling the costs and benefits of delayed product differentiation," *Management science,* vol. 43, no. 1, pp. 40-53, 1997.

[22] K. Lee and K. Joshi, "Information technology (IT) development in Korea: Past, present, and future," *Journal of Information Technology Case and Application Research,* vol. 17, no. 2, pp. 68-73, 2015.

[23] P. Lin and K. Saggi, "Product differentiation, process R&D, and the nature of market competition," *European Economic Review,* vol. 46, no. 1, pp. 201-211, 2002.

[24] A. Marcina and J. Erjavec, "SUPPLIER INVOLVEMENT IN NEW PRODUCT DEVELOPMENT," In *Proceedings of FEB Zagreb International Odyssey Conference on Economics and Business,* University of Zagreb, Faculty of Economics and Business, vol. 3, no. 1, pp. 901-914, 2021.

[25] D. M. Martínez and J. C. Fernández-Rodríguez, "Artificial intelligence applied to project success: a literature review," *IJIMAI,* vol. 3, no. 5, pp. 77-84, 2015.

[26] G. Marzi, "Beyond Pathologies in New Product Development," In *European Academy of Management 2021,* EURAM, pp. 1-20, 2021.

[27] K. J. Meagher and K. G. Zauner, "Product differentiation and location decisions under demand uncertainty," *Journal of Economic Theory,* vol. 117, no. 2, pp. 201-216, 2004.

[28] D. K. Nasiopoulos, D. P. Sakas, D. S. Vlachos and A. Mavrogianni, "Simulation of generation of new ideas for new product development and IT services," In AIP conference proceedings, American Institute of Physics, vol. 1644, no. 1, pp. 60-68, 2015.

[29] G. Nassimbeni and F. Battain, "Evaluation of supplier contribution to product development: fuzzy and neuro-fuzzy based approaches," *International Journal of Production Research,* vol. 41, no. 13, pp. 2933-2956, 2003.

[30] W. Oberhofer and J. Kmenta, "A general procedure for obtaining maximum likelihood estimates in generalized regression models," *Econometrica: Journal of the Econometric Society,* pp. 579-590, 1974.

[31] A. L. Page and G. R. Schirr, "Growth and development of a body of knowledge: 16 years of new product development research, 1989–2004," *Journal of Product Innovation Management,* vol. 25, no. 3, pp. 233-248, 2008.

[32] M. A. Raith, "Product differentiation, uncertainty and price coordination in oligopoly," *London School of Economics and Political Science (United Kingdom),* 1996.

[33] T. Rayna, "Understanding the challenges of the digital economy: the nature of digital goods," *Communications & Strategies,* vol. 71, pp. 13-26, 2008.

[34] S. Sabahi and M. M. Parast, "Firm innovation and supply chain resilience: a dynamic capability perspective," *International Journal of Logistics Research and Applications,* vol. 23, no. 3, pp. 254-269, 2020.

[35] Search.danawa.com. Accessed on: November 1, 2022. http://search.danawa.com/dsearch.php

[36] A. Shahin and S. Mohammadi Shahiverdi, "Estimating customer lifetime value for new product development based on the Kano model with a case study in automobile industry," *Benchmarking: An International Journal,* vol. 22, no. 5, pp. 857-873, 2015.

[37] W. R. Smith, "Product differentiation and market segmentation as alternative marketing strategies," *Journal of marketing,* vol. 21, no. 1, pp. 3-8, 1956.

[38] *Strategy Analytics*, Strategyanalytics.com, 2019. [Online]. Available: https://www.strategyanalytics.com/

[39] R. Tibishirani and K. Knight, "Model search and inference by bootstrap "bumping"," Technical Report, Department of Statistics, University of Toronto.(http://www-stat. stanford.-edu/tibs). Presented at the Joint Statistical Meetings, Chicago, 1996.

[40] M. Timko, "An experiment in continuous analysis," *Center for quality of Management Journal*, vol. *2,* no. 4, pp. 17-20, 1993.

[41] R. M. Van Steenbergen and M. R. Mes, "Forecasting demand profiles of new products," Decision support systems, vol. 139, 2020.

[42] *Versus*, Versus.com, 2022. [Online]. Available: https://versus.com/en

[43] Q. Wan, X. Zhao, H. Liu, H. Dinçer and S. Yüksel, "Assessing the New Product Development Process for the Industrial Decarbonization of Sustainable Economies," *SAGE Open*, vol. 12, no. 1, 2022.

[44] C. K. Williams and C. E. Rasmussen, "Gaussian processes for machine learning," *Cambridge, MA: MIT press,* vol. 2, no. 3, pp. 4, 2006.