THE STUDY OF PURCHASE INTENTION FOR MEN'S FACIAL CARE PRODUCTS WITH K-NEAREST NEIGHBOUR

Jui-Chan Huang *1, Pei-Yu Shao[2], Tzu-Jung Wu[3]

[1]Yango College, Fuzhou, 350015, China.
[2]Department of Business Management, National Kaohsiung Normal University, Kaohsiung, 80201, Taiwan.
[3]Institute of Human Resource Management, National Sun Yat-Sen University, Taiwan

*wish0718@outlook.com

ABSTRACT

Inventory management was a major issue for all the industries. The supplied of products to customers required the readiness of the inventory. This allowed rapid delivery and reduced waiting time for customers so that companies could profit from it. Any stock out or insufficiency would lead to loss of customers because their needs cannot be met. This would hurt firm profitability and market competitiveness. Inventory control was critical to retain liquidity and avoid overstocking. This was also the key to firm's survival and sustainability. To ensure an appropriate level of inventory, it was necessary to manage the inventory levels with sales forecast on an on-going basis. This paper tried to find out its inventory control in order to assisted Company T to improve its inventory control. Firstly, the products offered by Company T are classified into groups. The R programming language was then used to stimulate and forecast future sales of different products. Different techniques were applied to manage the inventory levels according to the results of categorizations and forecasts.; 3.Consolidation of all the product items and grouping them into activity-based classifications; 4.Simulation and forecasting of future sales according to the categorization results; 5. Formulation of different controlled techniques based on the simulations and forecasts, and application of these methods to inventory management.

KEYWORDS

Improvement of Inventory Control, Forecast, Activity-Based Classification

1. INTRODUCTION

A. Research Background

According to the statistical data of Industrial Development Bureau, Ministry of Economic Affairs (as of March 6, 2015), there were 680 cosmetics-related industries (including manufacturers, raw material suppliers, agents and others) throughout Taiwan as the biotechnological cosmetology developed rapidly in recent years. The Department of Statistics and Department of Commerce, Ministry of Economic Affairs calculated the sales volume of cosmetics and care products in Taiwan (from January 1982 to the end of December 2014). In the past five years, the sales volume of cosmetics and care products manifested an upward trend amidst dramatic volatility. In other words, the sales were rising. Moreover, with increasing personal income, people began to seek for higher quality of life and attached more and more importance to consumer products. Cosmetics and care products were no more luxuries beyond reach. Additionally, due to Taiwan's subtropical location, it was hot and humid in summer, and alternating dry, cold and damp in winter with a large temperature difference between day and night, which easily influenced the skin.

B. Research Motives

People, both men and women, had an inherent love for beauty. During the last few years, quite a few skin care product exclusive to men were found in all major sales channels of cosmetics and care products. The Internet, media, newspapers and magazines popularized the knowledge related to men's beauty care and released the experience in and advice on the use of care products. Therefore, it was not the exclusive right of women to apply and used care products any longer. Nevertheless, male users generally knew far less about care products, product efficacy and beauty care than women. When shopping, probably men would fear the chosen care.
Products might not be as effective as expected and thus hesitate about buying them. Hence, this study mainly explored men's purchase intention for facial care products.

C. Research Purposes

By the KNN algorithm commonly used in machine learning, this study aimed to build the classification model, reveal male consumers' purchased intention for facial care products, verify the influence on purchase intention and derive managerial implications from the data classification features.

2. LITERATURE REVIEW

This study inferred the factors that might influence male consumers' purchase intention for facial care products from the scholars' previous viewpoints on purchase intention and the literature about the influence on consumer purchase intention, and did a thorough literature review on the KNN algorithm verified by this study.

A. Definition of Purchase Intention

According to [1], purchase intention referred to consumers' knowledge and perception about products and their purchase patterns shaped by external environment. The strength of purchase intention could be considered as consumers' subjective feelings about the chosen products. Regarding purchase intention, Reference [2] pointed out that consumers made choices at the evaluation stage and decided the priority of purchase, based on which their purchase intention arose. Consumers would make decisions in line with their purchase intention. In their choice making process, other factors would affect their purchase intention as well. In the consumer behavior research of [3], purchase intention was used to measure the likelihood of product purchase.

B. Data Mining

Data mining, as part of knowledge-discovery in databases (KDD), was to search for useful patterns and relationships from the data by using many methods of statistical analysis and modeling. Reference [4] described the actions of data pre-processing: (1) cleansing; (2)

formatting (3) making up meaningful dialogues. Data mining was always related to computer science and reached the goals mentioned above in numerous ways, such as statistics, online analytic processing, information retrieval, machine learning, expert systems, past empirical rules and pattern recognition.

C. The functions of machine learning in data mining

According to [5], the main purposes of data mining included data patterns, description, relationships, classification, prediction, and clustering and evolution analysis. Data were used to build models; the models were applied to describe the patterns and relations in the data. After the models were built, they were useful in two ways: (1) the knowledge about data patterns and relations could provide the needed information when users made decisions; (2) data patterns could help users make predictions. Generally speaking, data mining had the following five functions: classification, estimation, prediction, affinity grouping and clustering, as described briefly in Table 1.

TABLE 1
FUNCTIONS OF DATA MINING

Mining	
Classification	Categorization was performed, definitions were given and categories were created in accordance with the attributes of the objects of analysis.
Estimation	The unknown value of a certain attribute was estimated according to the related attributes of existing continuous values
Prediction	The future value of the attribute was predicted based on the past observed values of the object attribute.
Affinity Grouping	It decided which related articles among all objects should be grouped together. This function was used to evaluate the chance of cross-selling and design attractive product groups.
Clustering	It differentiated the comparatively homogeneous clusters among the heterogeneous population. The data were differentiated without defining the differentiation in advance.

The methods of machine learning in data mining Reference [6] regarded machine learning as a technique of knowledge mining in data mining and the application of this technique to suitable tasks. Reference [7] divided the machine learning methods into two main categories, supervised learning algorithms and unsupervised learning algorithms, according to whether they had the training set or not. Table 2 briefly listed the commonly used methods in each category.
Literature review on related application of classification algorithms in data mining Reference [7] used the classification algorithm for knowledge mining and discovered the rule of relationships in a data set. The classification rule was generally described as X->Y; between X and Y was the cause-result relation. Reference [5] reviewed the processing knowledge discovery and data mining application in the manufacturing industry.

Classification was the learning to divide data items into several pre-defined categories. Usually, classification was performed in two steps. In the first step, a model was constructed to describe a set of predetermined data class or concepts. In this process, the training set was formed through the database tuple described by attribute analysis. In the second step, according to the classification precision, this model was used for the future classification of data or the testing set.

Table 2 Supervised Learning Algorithms

Supervised Learning Algorithms	
K-Nearest Neighbor	Classification
Naïve Bayes	Classification
Decision Trees	Classification
Classification Rule Learners	Classification
Linear Regression	Prediction
Regression Trees	Prediction
Model Trees	Prediction
Neural Networks	Classification, Prediction
Support Vector Machines	Classification, Prediction
Unsupervised Learning Algorithms	
Association Rules	Pattern Detection
K-means Clustering	Clustering

Reference [8] explored the dental implant failure cases in Taiwan based on their assumption that possible influence could be detected from the classification rule. The data of oral Implants were analyzed to generate the classification rule which could be used as a prediction method before the dental implant operation. Reference [10] regarded the rule of screening and classification as an important factor to improve the classification algorithm, and applied the classification algorithm to business intelligence, decision science and machine learning. More and more importance was attached to classification algorithms in recent years. The traditional classification algorithms often generated a lot of redundant and even contradictory information. Hence, to propose a new effective classification algorithm for big data set analysis became a key competitive and innovative element. The main challenge for an enterprise was how to use classification and became a key competitive and innovative element. The main challenge for an enterprise was how to use classification and applied it to big data. Dedicatedly developing and adopting compact rules and concurrently maintaining high classification accuracy was a new effective way to construct the classification rule.

Classification algorithm - K-Nearest Neighbor

For K-Nearest Neighbor (KNN) algorithm, K-Nearest Neighbor meant the K^{th} nearest neighbor. For example, there were 20 households living in this community. In other words, each household was your neighbor. The neighbor nearest to you was the "1st nearest neighbor"; the second nearest one was the "2nd nearest neighbor". The rest could be deduced in the same rule until the K^{th} one. From the perspective of [7[ ), the core idea of KNN algorithm was that the categories of the samples to be classified were determined only according to the categories of the nearest one or a few samples. Generally, the KNN classification was very suitable for classification tasks.

After using the KNN algorithm, Reference[11] presented a new rule which, improved based on the KNN method, allowed each object to belong to two given categories. Their research results showed the improved KNN efficacy actualized a deeper understanding of the data with less classification errors. Reference [12] found that it usually took many clients much time to get useful and correct information, so they combined the KNN algorithm in data mining with the system they developed themselves. As revealed by the results, the KNN classification featured transparency, consistency, ease of understanding and ideal qualities for a higher trend; moreover, it was easier to implement than other machine learning methods because there was no need to know the data allocation beforehand.

Table .3 Comparison of the Advantages and Disadvantages of KNN

Advantage	Disadvantage
Easier and more effective than other classification algorithms	A slow classification process
No need for any hypothesis about data allocation	Mass storage was needed to store data
Fast training	

Source:[7]

3. RESEARCH METHOD

A. Research Design

KNN Modeling The more knowledge consumers had about products and the better view they had on care, the stronger their purchase intention was. Otherwise, if the perceived risks could be reduced effectively, consumers would have a stronger purchase intention. Therefore, the average scores of "Product Perception" and "View on Care" on the questionnaire equal to or higher than 3 indicated a strong purchase intention; conversely, the average scores lower than 3 indicated a weak purchase intention. However, for "Perceived Risk", if the average score was equal to or lower than 3, it indicated a strong purchase intention; while the average score higher than 3 indicated a weak purchase intention. The purchase intention was taken as the output variable. The data went through the 3 indicated a weak purchase intention. The purchase intention was taken as the output variable. The data went through the min-max normalization and Z-score standardization so that all the eigenvalues ranged between 0 and 1. The top 75% of the data were set as "training" and the last 25% were set as "test" to construct the model. The class kit in the R language was applied to build the KNN model. For model evaluation, the cross tabulation was generated to compare the classification accuracy of these two different ways of standardization. The model with higher classification accuracy was chosen.

B. K-Nearest Neighbor

As mentioned in Chapter 2, [12] found that the KNN classification featured transparency, consistency, ease of understanding and ideal qualities for a higher trend. Therefore, this study used the KNN algorithm for modeling.

C. KNN-based distance calculation

There were many different ways to calculate the distance. But the most commonly used way of distance calculation in the KNN algorithm was Euclidean Distance, the straight-line distance between two points.

(1). Euclidean Distance equation:

$$d(x_i, x_j) = \sqrt{\left|x_{i1} - x_{j1}\right|^2 + \left|x_{i2} - x_{j2}\right|^2 + \dots + \left|x_{in} - x_{jn}\right|^2} \quad (3.1)$$

(2). Definition of the function:

x_{i1} : the 1st eigenvalue of data i.
x_{in} : the nth eigenvalue of data i.
x_{j1} : the 1st eigenvalue of data j.
x_{jn} : the nth eigenvalue of data j.

D. Appropriate K value

The method to determine the number of K values used in the model and converge the appropriate K value through the overtraining and undertraining data was referred to as "bias-variance tradeoff". Supposing a very large K value was set, similar to the observation of the extreme state of training data. In the opposite extreme situation, when the K value was set as 1, noise or an outlier would occur to the data, exerting adverse influence on classification. So obviously, the best K value ranged between these two extreme values. Generally speaking, the practical rule determined the K value to make it equal to the square root of the number of training data samples.

E. Data standardization

Before the use of the KNN algorithm, the eigen values in the data were generally transformed to a standard scope. The reason for this was that the distance formula depended on how to measure eigen values. Especially, if a certain eigen value was overly larger than other values, this distance measurement would be dominated greatly by the larger value. So, this value had to be standardized. The purpose of standardization was to make all the values in the data range between 0 and 1. The min-max normalization was generally performed for the KNN algorithm.

(1). Min-Max Normalization formula:

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (3.2)$$

(2). Definition of the function:

$x = (x_1, \dots, x_n)$, z_i was the ith normalized eigenvalue of in the data. x While another common transformation for the KNN algorithm was called Z-score Standardization, whose formula was shown as (3.9).

(3). Z-score Standardization formula:

$$z_i = \frac{x_i - \mu}{\sigma} \quad (3.3)$$

Definition of the function:

$x = (x_1, \ldots, x_n)$ and $\sigma \neq 0$.

μ : the mean of the population.

σ : the standard deviation of the population.

z_i : the ith standardized eigenvalue in the data.

4. Research Results Analysis And Discussion

A. KNN Classification

The R language was used for KNN modeling. Table 4 displayed the results of Min-Max Normalization for KNN classification. The three constructs in the questionnaire—Perceived Risk, Product Perception and View on Care—were used as input factors. The output factor was Purchase Intention. The first 75% of the data were taken as "training" and the last 25% as "testing". The K value, set as the square root of the training data (190), was approximately 14. The table showed the predicted values and actual outcomes and thus consisted of 4 areas. (1) Both the predicted and actual purchase intentions were weak. (2) Both the predicted and actual purchase intentions were strong. These two groups were the correct classification results with an accuracy of 87.50%. (3) The predicted purchase intention was weak while the actual purchase intention was strong. (4) The predicted purchase intention was strong while the actual purchase intention was weak. These two groups were the incorrect classification results with an error rate of 12.5%. As could be seen from the incorrect classification outcomes, the error rate for "predicted purchase intention strong but actual purchase intention weak" was much higher than that for "predicted purchase intention weak but actual purchase intention strong". Therefore, this study tried to improve the KNN model with Z-score Standardization.

Table.4 KNN Classification MIN-MAX Normalization

Min-Max / Actual	Predicted		Total
	Weak Purchase Intention	Strong Purchase Intention	
Weak Purchase Intention	2 (Person)	8 (Person)	10 (Person) 0.156 (10/64)
	0.200 (2/10)	0.800 (8/10)	
	1.000 (2/2)	0.129 (8/62)	
	0.031 (2/64)	0.125 (8/64)	
Strong Purchase Intention	0 (Person)	54 (Person)	54 (Person) 0.844(54/64)
	0.000 (0/54)	1.000 (54/54)	
	0.000 (0/2)	0.857 (54/62)	
	0.000 (0/64)	0.844 (54/64)	
Total	2 (Person)	62 (Person)	64 (Person) 1.00 (64/64)
	0.0031 (2/64)	0.969(62/64)	

Table 5 displayed the results of the KNN improvement with Z-score Standardization and also consisted of 4 areas according to the classification accuracy and error rate. (1) Both the predicted and actual purchase intentions were weak. (2) Both the predicted and actual purchase intentions were strong. In these two groups of correct classification, the accuracy dropped from 87.50% to 85.94%, mainly because the accuracy of the "both the predicted and actual purchase intentions were weak" group decreased. (3) The predicted purchase intention was weak while the actual purchase intention was strong. (4) The predicted purchase intention was strong while the actual purchase intention was weak. The error rate rose from 12.50% to 14.06%, mainly because the error rate of the "the predicted purchase intention was weak while the actual purchase intention was strong" group went up. Therefore, compared with Z-score Standardization, the originally-applied Min-Max Normalization achieved higher accuracy.

Table 5.KNN Classification Z-score Standardization

Z-score / Actual	Predicted		Total
	Weak Purchase Intention	Strong Purchase Intention	
Weak Purchase Intention	1 (Person)	9 (Person)	10 (Person) 0.156(10/64)
	0.100 (1/10)	0.900 (9/10)	
	1.000 (1/1)	0.143 (9/63)	
	0.016 (1/64)	0.141 (9/64)	
Strong Purchase Intention	0 (Person)	54 (Person)	54 (Person) 0.844(54/64)
	0.000 (0/54)	1.000 (54/54)	
	0.000 (0/1)	0.857 (54/63)	
	0.000 (0/64)	0.844 (54/64)	
Total	1 (Person)	63 (Person)	64 (Person) 1.00(64/64)
	0.016 (1/64)	0.984 (63/64)	

The practical rule to determine the K value made the K value equal to the square root of the training data. This study adopted the Min-Max Normalization for KNN Modeling. The K value, by the square root of 190, was approximately 13.78, so "14" was taken. Table 4.3 listed the comparison of different K values. Supposing the maximum K value was selected to observe the extreme classification reaction in the training data. When the K value was set as 1, the outlier emerged, exerting negative influence on classification. As could be seen from Table 8, when extreme K values-whether it be 1 or 64-were substituted, abnormal classification occurred with higher classification error rates in both cases. So, as the K value was set between these two extreme values, it could be found that the error rate was the lowest when K was 14 or 15, but it rose when a value higher than 15 was taken. Hence, this verified the practical rule that made the K value equal to the square root of the training data.

Table 5 Comparison of different K Values

#K Value	#Predicted-Weak But Actual-Strong	#Predicted-Strong But Actual-Weak	#Classification Error Rate
1	7	3	15.63%
5	6	3	14.06%
6	7	2	14.06%
12	9	0	14.06%
13	9	0	14.06%
14	8	0	12.50%
15	8	0	12.50%
16	9	0	14.06%
20	9	0	14.06%
25	9	0	14.06%
30	0	10	15.63%
40	0	10	15.63%

5. Conclusion And Managerial Implications

By using KNN for classification, this study explored whether perceived risk, product perception and view on care could influence the purchase intention. As hypothesized by this study, the purchase intention increased as consumers knew more about products and had a better view on care; otherwise, the purchase intention also became stronger if the consumer perceived risks were effectively reduced. The results showed that, the classification accuracy based on Min-Max Normalization was 87.5%, higher than 85.94%, the accuracy achieved by Z-score Standardization. Next, it was verified that the practical rule to make the K value equal to the square root of the training data achieved the best result. Therefore, the KNN algorithm could be actually used for classification. In this way, the research hypothesis was established: the purchase intention increased as consumers knew more about products and had a better view on care; otherwise, the purchase intention also became stronger if the consumer perceived risks were effectively reduced. As can be inferred from this pattern, it is advisable for manufacturers to mark in detail the steps to use products in their key stores, or further train their employees on how to use men's facial care products so that they could clear male consumers' doubts about facial care products or reassure, by giving proper answers, consumers about their desirable care products. Additionally, it was good to invite some famous bloggers on the Internet to post their feelings about using the products. This can help the men's facial care product users better understand the products and beauty care, reduce the risks of purchase, and strengthen their purchase intention.

REFERENCES:

[1] G.M., ullet, and M., arson,, Analysis of Purchase Intent Scales Weight by Probability of Actual Purchase, *Journal of Marketing Research*, 22(1), 93-96.1985.

[2] P. Kolter, , Marketing Management: Analysis, Planning, Implementation, & Control Englewood Cliffs, Prentice-Hall, Inc.: N. J.1991

[3] L. G., Schiffman, andL. L., Kanuk, , Consumer Behavior Upper Saddle River, *New Jersey Prent-Hall*.2000.

[4] Resul D.and T.C. Ibrahim, meaningful data from web logs for improving the impressiveness of a website by using path analysis method[J]Expert Systems with Applications, 36 (3), pp. 6635-6644. 2009.

[6] H. Jiawei, and K. Micheline, Data Mining Concept and Techniques(2nd), *Morgan Kaufmann Publishers*, 285–350, 2006.

[7] Lantz, B., Machine Learning with R, *Packt Publishing Ltd*.2013)

[8] Agrawal, R. et al, Mining Association Rules between Sets of Items in Large Databases, *Proceedings of the ACM SIGMOD Conference on Management of Data*, 207–216.1993.

[9] Chiang, H.J., Tseng, C.C., and Tomg, C.C, A Retrospective Analysis of Prognostic Indicators in Dental Implant Therapy Using The C5.0 Decision Tree Algorithm, *Journal of Dental Sciences*, 8, 248-255.2013.

[10] F. Chen, et al, , Principal Association Mining: An Efficient Classification Approach, Knowledge-Based Systems, 67, 16-25.2014.

[11] W.Y. Liu, et al, On Gender Differences in Consumer Behavior for Online Financial Transaction of Cosmetics, *Mathematical and Computer Modelling*, 58, 238-253.2013

[12] D.A., Adeniyi, Z.Wei,, and Y.Yongquan, Automated Web Usage Data Mining and Recommendation System Using K-Nearest Neighbor (KNN) Classification method,*Applied Computing and Informatics*, 12, 90-108.2015

[13] Aghekyan-Simonian, M. et al, The Role of Product Brand , 2012.