



# “Identifying customer priority for new products in target marketing: Using RFM model and TextRank”

<b>AUTHORS</b>	Seongbeom Hwang  Yuna Lee
<b>ARTICLE INFO</b>	Seongbeom Hwang and Yuna Lee (2021). Identifying customer priority for new products in target marketing: Using RFM model and TextRank. <i>Innovative Marketing</i> , 17(2), 125-136. doi: <a href="https://doi.org/10.21511/im.17(2).2021.12">10.21511/im.17(2).2021.12</a>
<b>DOI</b>	<a href="http://dx.doi.org/10.21511/im.17(2).2021.12">http://dx.doi.org/10.21511/im.17(2).2021.12</a>
<b>RELEASED ON</b>	Friday, 11 June 2021
<b>RECEIVED ON</b>	Monday, 19 April 2021
<b>ACCEPTED ON</b>	Thursday, 10 June 2021
<b>LICENSE</b>	 This work is licensed under a <a href="https://creativecommons.org/licenses/by/4.0/">Creative Commons Attribution 4.0 International License</a>
<b>JOURNAL</b>	"Innovative Marketing "
<b>ISSN PRINT</b>	1814-2427
<b>ISSN ONLINE</b>	1816-6326
<b>PUBLISHER</b>	LLC “Consulting Publishing Company “Business Perspectives”
<b>FOUNDER</b>	LLC “Consulting Publishing Company “Business Perspectives”



NUMBER OF REFERENCES

**44**



NUMBER OF FIGURES

**4**



NUMBER OF TABLES

**4**

© The author(s) 2021. This publication is an open access article.



**BUSINESS PERSPECTIVES**



LLC "CPC "Business Perspectives"  
Hryhorii Skovoroda lane, 10,  
Sumy, 40022, Ukraine  
[www.businessperspectives.org](http://www.businessperspectives.org)

**Received on:** 19<sup>th</sup> of April, 2021  
**Accepted on:** 10<sup>th</sup> of June, 2021  
**Published on:** 11<sup>th</sup> of June, 2021

© Seongbeom Hwang, Yuna Lee, 2021

Seongbeom Hwang, Specialist, Data Curation Team, LG Uplus Corp., South Korea. (Corresponding author)

Yuna Lee, Senior Officer, Digital Channel Unit, KB Securities co.Ltd., South Korea.

Seongbeom Hwang (South Korea), Yuna Lee (South Korea)

# IDENTIFYING CUSTOMER PRIORITY FOR NEW PRODUCTS IN TARGET MARKETING: USING RFM MODEL AND TEXTRANK

## Abstract

Target marketing is a key strategy used to increase the revenue. Among many methods that identify prospective customers, the recency, frequency, monetary value (RFM) model is considered the most accurate. However, no RFM study has focused on prospects for new product launches. This study addresses this gap by using website access data to identify prospects for new products, thereby extending RFM models to include website-specific weights. An RF model, built using frequency and recency information from website access data of customers, and an RwF model, built by adding website weights to frequency of access, were developed. A TextRank algorithm was used to analyze weights for each website based on the access frequency, thus defining the weights in the RwF model. South Korean mobile users' website access data between May 1 and July 31, 2020 were used to validate the models. Through a significant lift curve, the results indicate that the models are highly effective in prioritizing customers for target marketing of new products. In particular, the RwF model, reflecting website-specific weights, showed a customer response rate of more than 30% among the top 10% customers. The findings extend the RFM literature beyond purchase history and enable practitioners to find target customers without a purchase history.

## Keywords

RFM model, TextRank algorithm, new product, direct marketing, prospective customer, lift chart, telecommunications

## JEL Classification

C61, M31, M39

## INTRODUCTION

With the development of the Internet and the growth of online market, online consumption has increased, and recently COVID-19 has accelerated this trend (Bhatti et al., 2020). In this environment consumers can easily access and purchase products from multiple online stores. These trends increase the likelihood of being able to infer customers' interest from their online behavior. Analyzing their Internet access history allows to deduce certain items of interest and measure the likelihood of customers purchasing those items. Using this approach, companies that lack customers' purchasing histories can still efficiently identify marketing targets to launch their new products.

As the telecommunications market in the Republic of Korea is almost saturated, telecom companies are under pressure to cut rates and create new revenue sources (Yoon, 2007). New businesses must emerge to preserve or increase their profit. Telecommunications companies are well placed to develop hyperscale businesses that can identify customers' online behaviors. Specifically, they can gain insights into customer preferences with data generated from mobile phones and thereby, build new business that generate revenue (Chui & Manyika, 2015). Telecommunications companies want to launch new products rather



This is an Open Access article, distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Conflict of interest statement:**  
Author(s) reported no conflict of interest

than to rely on existing communications-related products, such as mobile phones and internet protocol TV. In this regard, a new product being launched is a content video service utilizing augmented reality/virtual reality technology. This service utilizes augmented reality to provide home training services such as yoga. However, it is important to break through early entry barriers because there are many operators providing home training content. One strategy to break into new markets is to identify targets and market to them directly (Stryker, 1996). However, it is not easy to identify marketing targets efficiently when entering a new business for the first time.

In target marketing, it is important to identify a customer base with a high probability of consumption to increase revenue. During this process, companies can use various types of data. One of the most common approaches is to conduct targeted marketing through recency, frequency, monetary value (RFM) analysis based on a purchase history of customers. However, this method can only be applied to businesses that have a portfolio of products and a history of purchases for each product. Therefore, it is difficult for companies that are new entrants to an existing market or pioneering a new business to identify prospective customers. It is important for a company to secure an initial market share. If the identification of targets and related activities are delayed compared to its competitors, this could result in a significant loss for the company.

## 1. LITERATURE REVIEW

### 1.1. RFM model

There are various ways to identify prospects and effectively conduct target marketing. Chou et al. (2000) stated that demographic information was used to define customer groups using the central value in a clustering technique (K-cluster, clique). Specifically, clusters were based on income data classified by customer occupation and gender (Chou et al., 2000). For verification, the results of the campaign execution by customer group were compared with a lift chart (Chou et al., 2000). This approach enables efficient identification of prospective customers using only demographic information (Chou et al., 2000). However, this approach has limitations as it presupposes that customers with the same demographic information have the same characteristics. Therefore, the most widely used approach is the RFM model.

Most studies commonly use a basic RFM model that assigns equal weights to recency (R), frequency (F), and monetary (M) value (Blattberg et al., 2008; Hughes, 1994; Khajvand et al., 2011; Nimbalkar & Shah, 2013). Additionally, Hosseini et al. (2010), Stone and Jacobs (1995), and Zaheri et al. (2012) conducted research to improve the target marketing accuracy of the traditional RFM model. One such improved model is the weighted RFM model, which assigns weights to each variable (i.e.,

R, F, and M). Studies using this model have suggested that weights should be assigned differently to R, F, and M depending on business characteristics (Stone & Jacobs, 1995). Thus, weights have been determined through verification by experts from each industry, along with the analytic hierarchy process technique, and this process has effectively addressed the limitations of the existing RFM model (Bin et al., 2008; Hosseini et al., 2010; Khajvand & Tarokh, 2011; Zaheri et al., 2012). Further, attempts have been made to increase the prospective customer discovery accuracy by adding new variables, in addition to R, F, and M (Hosseini et al., 2010; Wei et al., 2012). A key prospective group was found by proposing a length of RFM that applied to the purchase period in the RFM model (Wei et al., 2012). Another proposed method has been to segment customer groups using K-means clustering after adding the period of product activity as a variable to the RFM model. A verification of this study was conducted using data from an Iranian delivery company, SAPCO (Hosseini et al., 2010). As a result of the verification, the proposed model showed a significant effect on segmenting customer groups based on loyalty (Hosseini et al., 2010). In a separate study, customer groups with similar purchasing behaviors were classified using purchase data and a genetic-algorithm-based clustering technique (Tsai & Chiu, 2004). Then, an RFM model was used for analysis between clusters (Tsai & Chiu, 2004). The study proposed an RFM model for each product

(RFM/P), considering the product perspective rather than that of a customer. This model first estimated the customer value of all products individually and then aggregated them to obtain the overall customer value. Based on this, Heldt et al. (2019) conducted an empirical analysis on financial companies and supermarkets and the authors were able to verify the significance of results.

## 1.2. Application of the RFM model

The RFM model measures when, how often, and how much a customer purchases a product. History of customer purchases can effectively predict future purchasing behavior, and companies can estimate the value of the customers that contribute to their revenue. Thus, RFM models are widely applied in database marketing. For example, they are often developed for marketing programs (i.e., direct mail) targeting specific customers to improve customer response rates (Sohrabi & Khanlari, 2007). This shows that RFM facilitates the choice which customers to target through proposals (Colombo & Jiang, 1999).

Companies can benefit from the adoption of RFM models in terms of increased response rates, reduced order costs, and increased profits. In many studies, RFM models based on purchase history data from various industries have been applied to define high-value customers, who are the most attractive potential customers, and real-world data have verified the significance of high-value customers. The RFM model is used in different fields, such as finance (Abbasimehr & Shabani, 2020; Heldt et al., 2019; Hsieh, 2004), mobile communication (Bacila et al., 2012; Babaiyan & Sarfarazi, 2019; Li et al., 2008; Zabkowski, 2016), government (Bin et al., 2008), retail (Aggarwal & Yadav, 2020; Dogan et al., 2018; Heldt et al., 2019; Rizki et al., 2020; Wu et al., 2020), hospitals (Wei et al., 2016), and travelling (Lumsden et al., 2008). Although different industries have been analyzed, these studies identify high-value customers to shed light on topics such as loyalty, purchase, and revisit behaviors through RFM models. RFM models have been proved to be effective in many industries as a means of finding target customers.

The specific aspect of new product launch has not been considered before, including the studies dis-

cussed above. In most studies, the main contribution is refining the existing RFM models. Though, theoretically, upgrading a model is a positive improvement, as introducing new aspects is also necessary. Therefore, a method for the effective application of the RFM model in the context of new product launches (without any purchasing history) is proposed. This aspect is a concern for all practitioners who seek target consumer data without a purchase history, and this research aims to address this concern.

## 2. AIMS AND HYPOTHESES

### 2.1. Aims

This study proposes a method to prioritize valid marketing targets through website access for related product categories when there is no product purchase history. It also establishes a hypothesis that the more frequently customers access a website, the more priority they should have in target marketing. This hypothesis and the methodologies developed are subsequently verified through real data. Thus, this research aims to guide companies seeking marketing targets for the launch of new products.

### 2.2. Hypotheses

This study applies an RFM model that assumes shopping website access represents a history of purchases. Further, it defines hypotheses to see if it is meaningful to use website access history to predict customer value. It also considers access time and number of connections because website access history does not contain purchase amount data, unlike purchase history data. In particular, many websites may be accessed in relation to a product or service, but because of website differences, such as quality (Griffiths & Christensen, 2005) and quantity, customers tend to access certain websites frequently. This has implications for the research because among websites related to product categories, which are frequented by customers, and some have higher utility weights than others for research purposes. Thus, customers that are more likely to access these sites with higher utility weight will demonstrate more positive buying behavior.

Accordingly, based on the aims and the literature review, the following three hypotheses are proposed:

- H1: *Using website access data related to the product categories in the RFM model, instead of historical purchase data, leads to significant target marketing results.*
- H2: *There are websites related to product categories that several customers access frequently.*
- H3: *In target marketing, the RFM model, which weighs websites that are frequently accessed by people, is more meaningful than the model that does not take weight into account.*

### 3. METHODOLOGY

#### 3.1. Research model

Figure 1 illustrates the conceptual RFM model for identifying prospective customers adopted in this study. As mentioned in the literature review, this model is adopted because it has a strong ability to find potential customers. In addition, this study expands this model to identify potential customers for new products even without having a purchase history. Therefore, data on the purchase history of the product is replaced with the website access history related to product categories.

Accordingly, the RFM model is redefined as the RF model because monetary information cannot be defined within the website access history.

Then, the RwF model, which reflects website-specific utility weights, and the RF model, which does not, are applied to analyze prospects and thereby to evaluate the performance of the two models through empirical research.

The research analysis process comprised the following steps.

Step 1: Data pre-processing. Initially, web access history data related to product categories were extracted. Then, outliers and inaccurate values were removed and the initial dataset was generated. Next, redundant properties were removed to transform the data into an easier and more efficient form for processing to analyze customer value.

Step 2: Define factors for the model that identify prospects. The factors of the RF and RwF (*recency, frequency, and utility-weighted frequency*) are defined as follows.

Recency: The last website access shown by R represents the most recent value in website access and is defined as a gap between the start date of a given period and the date of website access. It is normalized to a value between 0 and 1. The higher the recency, the greater the R.

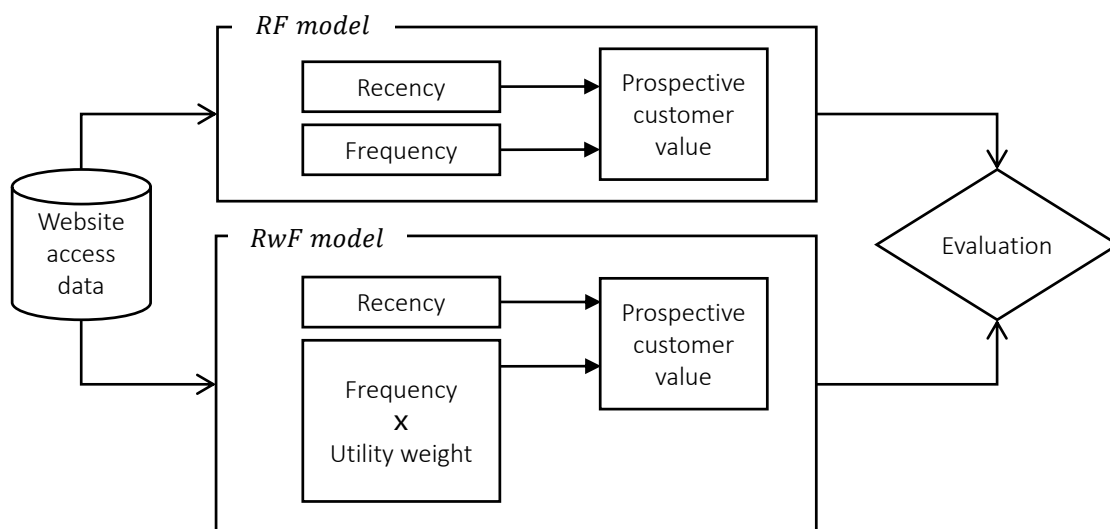
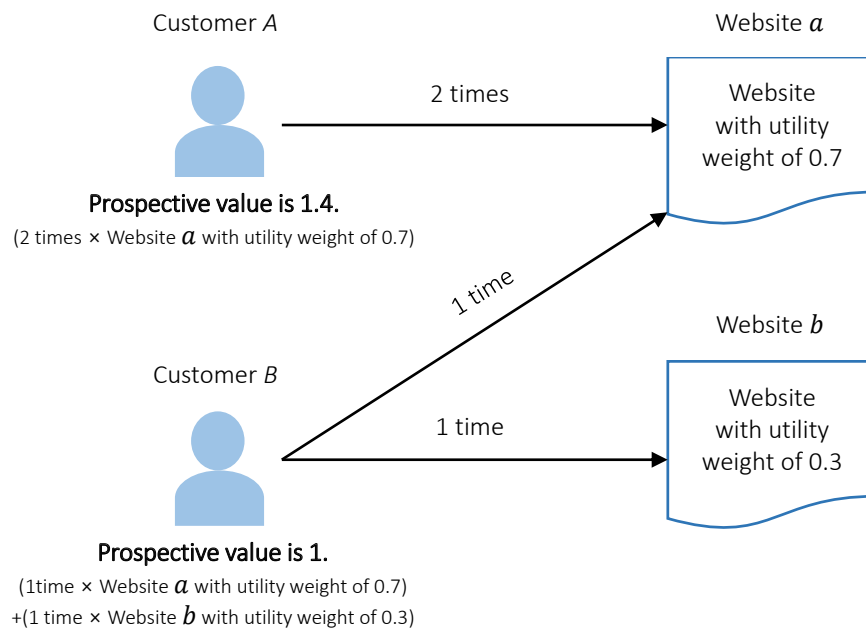


Figure 1. Conceptual framework to identify prospective customers



**Figure 2.** Example of the difference in the website pages accessed by customers

Frequency: This refers to the number of website visits to the same site during a given period. This number is normalized to a value between 0 and 1. The greater the frequency of visits, the greater the F.

Utility-weighted frequency: This refers to the multiplication of utility weights and access frequency by website. Each website offers different degrees of purchasing opportunities to customers. As each website has different types, quantities, and quality of information, each website has different utility weights. As shown in Figure 2, supposing there are websites and associated with a specific category, such as yoga, and website has a utility weight of 0.7, website has a utility weight of 0.3. Consumer accesses website twice; therefore, the weighted sum of website utility and number of times accessed is 1.4 ( $= 0.7 \text{ weight} \times 2 \text{ times}$ ). Consumer accesses website once and website once, and the weighted sum of website utility and the number of times accessed becomes 1 ( $= 0.7 \text{ weight} \times 1 \text{ time} + 0.3 \text{ weight} \times 1 \text{ time}$ ). In terms of website access frequency, consumers and have the same two connections. However, in terms of the weighted value of the website utility and number of connections, customer is 1.4 and customer B is 1. Thus,

customer has higher value than customer. Even if consumers access the website the same number of times, the weighted sum value can vary depending on the specific website pages they access.

To quantify the utility weights for each website page, as shown in Figure 2, the study utilizes the simultaneous access frequency of customers. Websites that attract many customers are defined as having high utility weights. A TextRank algorithm is used for analysis of these website-specific utility weights based on the frequency of the customer access of each website. The frequency term of the model is defined by the weighted sum of utility weights and access counts for each website. The RF model and the Rwf model differ in the frequency term, whereby the latter reflects the utility weights for each website, not just the frequency itself.

The PageRank algorithm places relative importance on documents with hyperlink structures (Ding et al., 2009). This algorithm has been used in many studies in various fields to find the key indicator of website importance (Heddiemann et al., 2010; Wang et al., 2013; Yan & Ding, 2011). A study of social network analyses looking for key users



found that the PageRank algorithm resulted in better results than traditional approaches, such as degree centrality and closing centrality (Mihalcea & Tarau, 2004). Griffiths and Christensen (2005) used PageRank as an indicator of the quality of web pages as experienced by consumers. The PageRank algorithm is used in various fields and can even be applied to find keywords in text analysis. A sentence is split into words to create a co-occurrence matrix for each word. The weight for each word is calculated using the TextRank algorithm, which is applied by the PageRank algorithm for text analysis. A word with a high TextRank value is considered a keyword for that text (Jaffery & Liu, 2009; Tian, 2013).

The method to obtain website-specific weights is a TextRank algorithm that finds keywords through word-specific weight calculations (Tian, 2013). To apply the TextRank algorithm to text, a non-directional graph is defined, in which words are expressed as a link between words as nodes. Applying these TextRank algorithms, the TextRank formula for each website is defined as

$$TR(Web_i) = (1-d) + d \cdot \sum_{Web_j \in In(Web_i)} \frac{w_{ij}}{\sum_{Web_k \in Out(Web_j)} w_{jk}} TR(Web_j) \tag{1}$$

$TR(Web_i)$  is the TextRank value by website, and  $w_{ij}$  is the co-occurrence value between websites. The link value between node  $Web_i$  and  $Web_j$  can be expressed as  $w_{ij}$ . The total sum of  $TR(Web_i)$  will be 1. The damping factor  $d$  is usually set to 0.85.

The steps for analyzing  $TR(Web_i)$  by website are as follows:

1. Create a customer-website matrix ( $m$ ). The value of the matrix is the number of times a customer accesses each website.
2. Create a co-occurrence matrix ( $m^T m$ ) between websites.
3. Analyze the weight value of each web page through the TextRank algorithm.

Using these steps, the weighted value of each website is determined, and it is higher for websites that

are frequently accessed by a relatively large number of customers. Then, the TextRank algorithm is used as the utility weight to define the weight for each website.

Step 3: Normalization of factors in models. Given the differences in scale, the data for the three factors of the models – namely, recency, frequency, and utility-weighted frequency – need to be standardized. Obtaining initial standardization using min-max standardization methods eliminates the impact of the respective numerical values on the analytical results.

$$x_{ij} = \frac{x_{ij} - \min\{x_j\}}{\max\{x_j\} - \min\{x_j\}} \tag{2}$$

Step 4: Formulate and analyze identifying prospective models. The RF and Rwf values are analyzed for each prospective customer according to the following expressions.

When applied as defined above, the RF model is shown as

$$RF_c = R_c + \sum_{s=1}^n F_{c,i} \tag{3}$$

where  $R_c$  is the recency value indicating the recency of the date of website access for customer  $c$ , and  $F_{c,i}$  is the total access frequency of the  $i$  website in a given period of time for customer  $c$ .

The Rwf model that reflects the utility weight by websites is as follows:

$$RwF_c = R_c + \sum_{i=1}^n F_{c,i} \cdot TR(Web_i) \tag{4}$$

where  $TR(Web_i)$  is a TextRank value based on the co-occurrence of the  $i$  website. Both the  $RF$  and  $RwF$  model results are normalized between 0 and 1.

Step 5: Evaluate. As target marketing aims to find prospective customers with a high value for purchases rather than to match all prospective purchase consumers accurately, a lift chart on a quantitative scale can be used to meet this objective (Jaffery & Liu, 2009; Piatetsky-Shapiro & Masand, 1999). This lift chart is often used in studies that measure the effectiveness of target marketing

(Burez & Van den Poel, 2007; Chou et al., 2000; Kim & Street, 2004). The equation of the lift chart for each upper percentile is as follows:

$$lift_p = \frac{Hits_p / Record_p}{Total Hits / Total Record} \quad (5)$$

$Hits_p$  is the number of successful marketing target consumers up to the  $p$ -percentile, and  $Record_p$  is the number of marketing targets up to the  $p$ -percentile. The percentile of the lift chart usually measures the performance by dividing it into 10% units.

To determine the significance of the model used in the study, the mobile long message service (LMS) is used to contact the top N% prospective value customers, and then the service subscription rate is compared. The LMS content consists of a description of the service and a subscription URL. The LMS campaign targets are sampled from 1,000 people for each 10% section of values analyzed from each RwF and RF model. In addition, 10,000 people are sampled for each RF and RwF model so that the targets of the campaign do not overlap. Then, LMS campaigns are conducted for these customers.

## 4. RESULTS

### 4.1. Data collection

Website access data are collected from a mobile telecommunications company in the Republic of Korea. The study collects and uses the history data of website access for 357,856 mobile subscribers from May 1 to July 31, 2020. The study focusses on a home training service with augmented reality, which is scheduled to be launched, and which will offer various elements of content, such as yoga and fitness, in the future. However, as most of the content at the time of launch is related to yoga categories, history data of website access related only to yoga is used for this study. The list of websites for the study include related yoga category websites, which are mainly comprised of yoga shopping offerings and yoga academy websites where consumption can occur. A list of 156 items had been collected by the mobile communication company on its own and verified manually. The access histo-

ry dataset for yoga websites consists of the customer ID, website name, website category, number of connections, and access date (see Table 1).

**Table 1.** Partial website access data

Customer ID	Website URL	URL category	Access count	Year-Month-Day
AKSADDK	www.xexymix.com	Yoga	12	2020-05-01
AKSADDK	www.xexymix.com	Yoga	5	2020-05-23
AKSADDK	www.healthyfriends.co.kr	Yoga	2	2020-06-05
SADJLSAD	www.andar.co.kr	Yoga	6	2020-07-01
WQIQIWEJ	ashtangayoga.kr	Yoga	2	2020-05-12
DSASAWEE	eadenyoga.modoo.at	Yoga	2	2020-05-28
...	...	...	...	...

Note: ID = identification; URL = uniform resource locator.

### 4.2. Prospective customer value by models

The statistics (average, maximum, and minimum) of the prospective customer values calculated with the RF and RwF models is shown in Table 2 and Table 3. The statistics for each decimal interval is distributed without overlapping intervals. This shows that each decimal segment has a different prospective purchasing power intensity. The statistics shows that the values calculated from the RF model are higher than those from the RwF model. This is due to the fact that there is no difference in weight between websites.

**Table 2.** RF value statistics

Decile	Number of customers	RF value		
		Mean	Max	Min
1	35,785	0.411	1.000	0.331
2	35,785	0.352	0.512	0.229
3	35,785	0.295	0.311	0.195
4	35,785	0.217	0.288	0.176
5	35,786	0.183	0.211	0.131
6	35,786	0.154	0.195	0.117
7	35,786	0.131	0.177	0.097
8	35,786	0.118	0.151	0.058
9	35,786	0.079	0.123	0.033
10	35,786	0.051	0.072	0.000

Note: R = recency; F = frequency.



**Table 3.** RwF values statistics

Decile	Number of customers	RwF value		
		Mean	Max	Min
1	35,785	0.398	1.000	0.296
2	35,785	0.255	0.296	0.225
3	35,785	0.203	0.225	0.185
4	35,785	0.167	0.185	0.151
5	35,785	0.137	0.151	0.121
6	35,785	0.105	0.121	0.089
7	35,787	0.077	0.089	0.065
8	35,786	0.050	0.065	0.037
9	35,786	0.030	0.037	0.021
10	35,787	0.012	0.021	0.000

Note: *R* = recency; *wF* = utility-weighted frequency.

### 4.3. Estimation results for target marketing

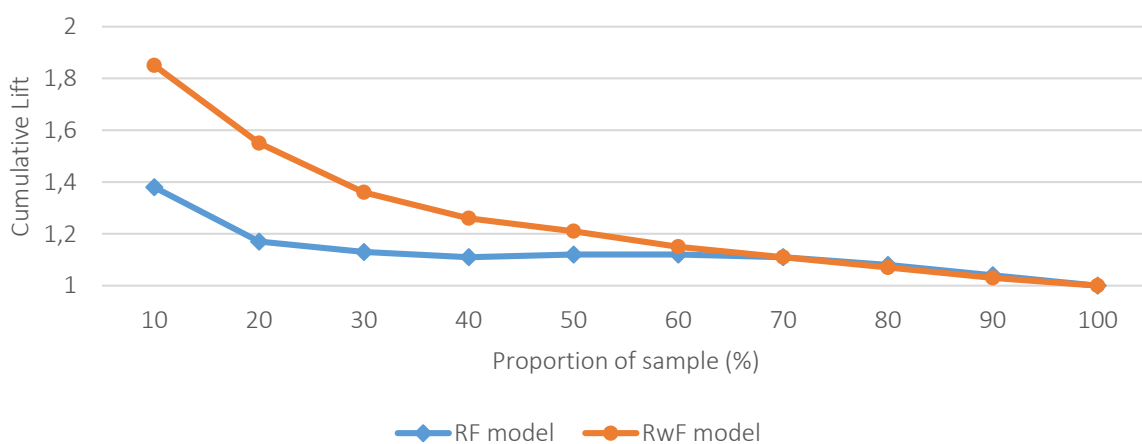
As a result, the total number of customers, which are expected to subscribe, is 618 for the RF model and 623 for the RwF model. Table 4 shows that the lift value of the RF model represents an overall decreasing trend. According to the lift value for each percentile segment, there is a slight increase in the lift value in the 50<sup>th</sup> percentile section, but it does not have enough influence to reverse the declining trend. In comparison, the lift value of the RwF model continues to decrease from the 10<sup>th</sup> percentile to the 100<sup>th</sup> percentile. General practice suggests that the higher the interval, the higher the lift value is, so it can be argued that both models are significant for target marketing. These results support H1.

As mentioned previously, high TextRank websites are those with high utility weights. The RwF model that reflects this utility is found to be 34% higher

than the RF model in the 10<sup>th</sup> percentile section, and the 20<sup>th</sup> percentile interval is 32% higher. Up to the 50<sup>th</sup> percentile section, it can be confirmed that the RwF model represents a higher lift value than the RF model. Moreover, the 80<sup>th</sup> percentile interval indicates that the RF model is 1% higher than the RwF model. Typically, if the slope of a lift chart decreases sharply to the right, the discovery of the top prospective customers has been performed well (Wang et al., 2013). Figure 3 shows that the RwF model, which has a large difference in slope of the lift value up to the 30<sup>th</sup> percentile section, is more effective in target marketing than the RF model. For this reason, H3 is supported only for highly prospective customers.

### 4.4. Estimation results for website utility weight

The website cumulative weight distribution represents 80% of the total weight of the top-40 web-



**Figure 3.** Cumulative lift chart of the RwF and RF models (*R* = recency; *w* = utility-weighted frequency; *F* = frequency)

**Table 4.** Performance results of the models

Proportion of sample (%)	RF model		RwF model	
	Cumulative hits	Cumulative lift	Cumulative hits	Cumulative lift
10	85	1.38	115	1.85
20	145	1.17	193	1.55
30	210	1.13	254	1.36
40	275	1.11	314	1.26
50	347	1.12	376	1.21
60	415	1.12	431	1.15
70	481	1.11	482	1.11
80	533	1.08	532	1.07
90	578	1.04	579	1.03
100	618	1.00	623	1.00

Note: R = recency; w = utility-weighted frequency; F = frequency.

sites, as shown in Figure 4. In particular, the top-10 websites account for 70% of the total weight, which implies that these are the websites commonly accessed by consumers interested in yoga. The top website has a weight of approximately 0.2, meaning that it has a huge influence in the field of yoga. In other words, there are websites that are relatively frequently accessed by customers and those that are not. Therefore, H2 is strongly supported.

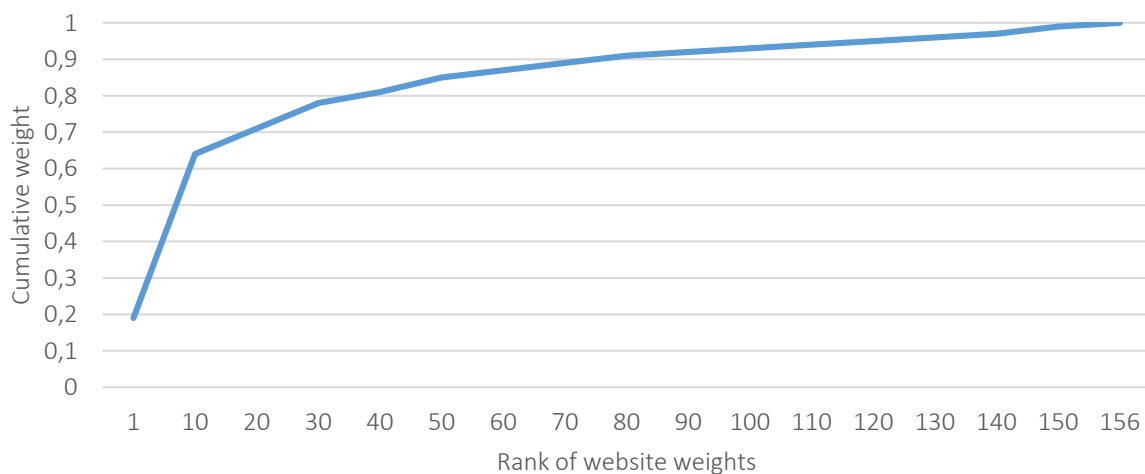
## 5. DISCUSSION AND IMPLICATIONS

Similarly, to Babaiyan and Sarfarazi (2019), Heldt et al. (2019), and Wu et al. (2020), RFM models were applied and expanded to further confirm that the RFM model is a powerful tool for finding prospective customers. However, the results of this study reflect two differences. First, the possibility of utilizing website

access history instead of purchase history was verified. Second, the RwF model, which provided different prospects depending on which website customers accessed, was significantly better in identifying prospects than the RF model.

Both RF and RwF models proposed were tested with LMS campaigns and represented higher target marketing success rates for customers with higher prospects. This means that access to websites for related product categories is a positive factor affecting product purchase. It also confirmed that using a website access history could be an important approach to prioritizing potential customers if a new product launch company or marketer does not have a consumer purchase history.

The TextRank algorithm was used to identify websites, which should be more heavily weighted than others. Simply making assumptions based on the number of website pages visited



**Figure 4.** Cumulative distribution of website weight

can be ineffective because it may reflect the number of connections generated by a small number of customers. However, the TextRank method, based on the co-occurrence matrix between websites, allocates higher weights to sites that consumers commonly access. This TextRank approach can overcome the problem of finding representative websites simply by the number of connections. This meant that most prospective customers could be identified to access the top 25% of websites, as determined by the website utility weights.

This study proposed an RwF model with utility weights by website and confirmed that the actual LMS campaign, which was based on the RwF model, showed a higher lift value than the RF model. In particular, the value of the lift chart in the top 10% section was 23% higher than that of the RF model, and the difference between

the top 50% section was more than 5%. This shows that customers are more likely to sign up for products on websites that they access more frequently. This approach will improve the implementation of higher-accuracy target marketing for companies launching new products. As consumption-related activities, such as information exploration and purchasing, gradually shift to online-based activities, the value of the approach developed in this study is likely to become an ever more powerful way to estimate individual interests and discover prospective customers.

Specifically, telecommunications companies or firms, which collect website access data, can apply study results immediately. The proposed approach can be effective in prioritizing prospective customers in the early stages of the launch of various products.

---

## CONCLUSIONS AND FUTURE WORK

This study sought to develop a procedure to identify prospective customers by priority for targeted marketing for new product launches. Notably, the model yields the value of the customer likely to be interested in a new product. The primary output of the proposed procedure is a decision-making process that enables an entity entering a new business to identify easily customers to be targeted with new offerings. That is, the model identifies which customers are more important, and thus are likely to contribute more to firm revenue. The model allows marketing practitioners to analyze the value of marketing targets, prioritize them, and market to as many customers as necessary, even if they do not have a purchase history.

Ultimately, the proposed approach could be used by a variety of companies seeking to launch new products. Future research is required to further verify the performance of the model by applying it to products in various fields. Collecting access time data that detail how long a customer has spent on a website will help to find prospective customers more accurately through their application to the website weight analysis.

## AUTHOR CONTRIBUTIONS

Conceptualization: Seongbeom Hwang, Yuna Lee.

Data curation: Seongbeom Hwang.

Formal analysis: Seongbeom Hwang.

Investigation: Seongbeom Hwang, Yuna Lee.

Methodology: Seongbeom Hwang.

Software: Seongbeom Hwang.

Supervision: Seongbeom Hwang.

Validation: Seongbeom Hwang.

Visualization: Seongbeom Hwang, Yuna Lee.

Writing – original draft: Seongbeom Hwang.

Writing – review & editing: Seongbeom Hwang, Yuna Lee.

## REFERENCES

1. Abbasimehr, H., & Shabani, M. (2020). A new framework for predicting customer behavior in terms of RFM by considering the temporal aspect based on time series techniques. *Journal of Ambient Intelligence and Humanized Computing*, 12, 515-531. <https://doi.org/10.1007/s12652-020-02015-w>
2. Aggarwal, A. G., & Yadav, S. (2020). Customer segmentation using Fuzzy-AHP and RFM model. *2020 8<sup>th</sup> International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*, 77-80. <https://doi.org/10.1109/ICRITO48877.2020.9197903>
3. Babaiyan, V., & Sarfarazi, S. A. (2019). Analyzing customers of South Khorasan telecommunication company with expansion of RFM to LRFM model. *Journal of AI and Data Mining*, 7(2), 331-340. <https://dx.doi.org/10.22044/jadm.2018.6035.1715>
4. Bacila, M. F., Radulescu, A., & Marar, I. L. (2012). RFM based segmentation: an analysis of a telecom company's customers. *The Proceedings of the International Conference "Marketing – from Information to Decision"*. Retrieved from <https://www.proquest.com/openview/88b949e5655be50727454b504a942986/1?pq-origsite=gscholar&cbl=1606336>
5. Bhatti, A., Akram, H., Basit, H. M., Khan, A. U., Naqvi, S. M. R., & Bilal, M. (2020). E-commerce trends during COVID-19 pandemic. *International Journal of Future Generation Communication and Networking*, 13(2), 1449-1452. Retrieved from <https://lovackey.eu/ws/media-library/8291b8bb61d0458d9bec753432dc4842/qualitative-paper.pdf>
6. Bin, D., Peiji, S., & Dan, Z. (2008). Data mining for needy students identify based on improved RFM model: a case study of university. *Proceedings of the 2008 International Conference on Information Management, Innovation Management, and Industrial Engineering*, 1, 244-247. <https://doi.org/10.1109/ICIMI.2008.128>
7. Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2008). RFM analysis. *Database Marketing*, 18, 323-337. [https://doi.org/10.1007/978-0-387-72579-6\\_12](https://doi.org/10.1007/978-0-387-72579-6_12)
8. Burez, J., & Van den Poel, D. (2007). CRM at a pay-TV company: using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 32(2), 277-288. <https://doi.org/10.1016/j.eswa.2005.11.037>
9. Chou, P. B., Grossman, E., Gunopulos, D., & Kamesam, P. (2000). Identifying prospective customers, in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 447-456. <https://doi.org/10.1145/347090.347183>
10. Chui, M., & Manyika, J. (2015). *Competition at the digital edge: 'Hyperscale' businesses*. The McKinsey Quarterly. Retrieved from <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/competition-at-the-digital-edge-hyperscale-businesses>
11. Colombo, R., & Jiang, W. (1999). A stochastic RFM model. *Journal of Interactive Marketing*, 13(3), 2-12. [https://doi.org/10.1002/\(SICI\)1520-6653\(199922\)13:3%3C2::AID-DIR1%3E3.0.CO;2-H](https://doi.org/10.1002/(SICI)1520-6653(199922)13:3%3C2::AID-DIR1%3E3.0.CO;2-H)
12. Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60(11), 2229-2243. <https://doi.org/10.1002/asi.21171>
13. Dogan, O., Ayçin, E., & Bulut, Z. (2018). Customer segmentation by using RFM model and clustering methods: a case study in retail industry. *International Journal of Contemporary Economics and Administrative Sciences*, 8(1), 1-19. Retrieved from <https://www.semanticscholar.org/paper/CUSTOMER-SEGMENTATION-BY-USING-RFM-MODEL-AND-A-CASE-Dogan-Ay%C3%A7in/11f8e998bdd759a8c576bad4e1cd66388a168920>
14. Griffiths, K. M., & Christensen, H. (2005). Website quality indicators for consumers. *Journal of Medical Internet Research*, 7. Retrieved from <https://www.semanticscholar.org/paper/Website-Quality-Indicators-for-Consumers-Griffiths-Christensen/459891f9dd2f017a43d0daf63c5c4ea9a675564d>
15. Hediemann, J., Klier, M., & Probst, F. (2010). Identifying key users in online social networks: a PageRank based approach. *Proceedings of the 31st International Conference on Information Systems*. Retrieved from <https://core.ac.uk/display/11554218>
16. Heldt, R., Silveira, C. S., & Luce, F. B. (2019). Predicting customer value per product: from RFM to RFM/P. *Journal of Business Research*, 127, 444-453. <https://doi.org/10.1016/j.jbusres.2019.05.001>
17. Hosseini, S. M. S., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, 37(7), 5259-5264. <https://doi.org/10.1016/j.eswa.2009.12.070>
18. Hsieh, N. C. (2004). An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert Systems with Applications*, 27(4), 623-633. <https://doi.org/10.1016/j.eswa.2004.06.007>
19. Hughes, A. M. (1994). *Strategic database marketing: the master plan for starting and managing a profitable, customer-based marketing program*. Chicago, IL: Probus Publishing.
20. Jaffery, T., & Liu, S. X. (2009). *Measuring campaign performance by using cumulative gain and lift chart* (Paper 196-2009). S.A.S. Global Forum. Retrieved from <http://support.sas.com/resources/papers/proceedings09/196-2009.pdf>
21. Khajvand, M., & Tarokh, M. J. (2011). Estimating customer future value of different customer segments based on adapted RFM model in retail banking context. *Procedia Computer Science*, 3, 1327-1332. <https://doi.org/10.1016/j.procs.2011.01.011>

22. Khajvand, M., Zolfaghar, K., Ashoori, S., & Alizadeh, S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behavior: case study. *Procedia Computer Science*, 3, 57-63. <https://doi.org/10.1016/j.procs.2010.12.011>
23. Kim, Y., & Street, W. N. (2004). An intelligent system for customer targeting: a data mining approach. *Decision Support Systems*, 37(2), 215-228. [https://doi.org/10.1016/S0167-9236\(03\)00008-3](https://doi.org/10.1016/S0167-9236(03)00008-3)
24. Li, S. T., Shue, L. Y., & Lee, S. F. (2008). Business intelligence approach to supporting strategy-making of ISP service management. *Expert Systems with Applications*, 35(3), 739-754. <https://doi.org/10.1016/j.eswa.2007.07.049>
25. Lumsden, S. A., Beldona, S., & Morrison, A. M. (2008). Customer value in an all-inclusive travel vacation club: an application of the RFM framework. *Journal of Hospitality and Leisure Marketing*, 16(3), 270-285. <https://doi.org/10.1080/10507050801946858>
26. Miglautsch, J. R. (2000). Thoughts on RFM scoring. *Journal of Database Marketing & Customer Strategy Management*, 8(1), 67-72. <https://doi.org/10.1057/palgrave.jdm.3240019>
27. Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into text. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 404-411. Retrieved from <https://www.aclweb.org/anthology/W04-3252/>
28. Nimbalkar, D. D., & Shah, P. (2013). Data mining using RFM analysis. *International Journal of Scientific and Engineering Research*, 4(12), 940-943. Retrieved from <https://www.ijser.org/researchpaper/Data-mining-using-RFM-Analysis.pdf>
29. Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank Citation Ranking: Bringing Order to the Web* (Technical Report). Stanford InfoLab. Retrieved from <http://ilpubs.stanford.edu:8090/422/>
30. Piatetsky-Shapiro, G., & Masand, B. (1999). Estimating campaign benefits and modeling lift. *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 185-193. Retrieved from <https://www.kdnuggets.com/gp-spubs/kdd99-estimating-campaign-benefits-modeling-lift.pdf>
31. Rizki, B., Ginasta, N. G., Tamin, M. A., & Rahman, A. (2020). Customer loyalty segmentation on point of sale system using recency-frequency-monetary (RFM) and K-means. *Jurnal Online Informatika*, 5(2), 130-136. Retrieved from <https://join.if.uinsgd.ac.id/index.php/join/article/view/511>
32. Sohrabi, B., & Khanlari, A. (2007). Customer lifetime value (CLV) measurement based on RFM model. *Iranian Accounting and Auditing Review*, 14, 7-20. Retrieved from [https://acctgrev.ut.ac.ir/article\\_18552\\_3c49bb9a53ce8058c5e7d811b5515f2d.pdf](https://acctgrev.ut.ac.ir/article_18552_3c49bb9a53ce8058c5e7d811b5515f2d.pdf)
33. Stone, B., & Jacobs, R. (1995). *Successful Direct Marketing Methods*. Lincolnwood, IL: NTC Business Books.
34. Stryker, J. D. (1996). Launching a new business-to-business product. In M. D. Rosenau Jr., A. Griffin, G. Castellion, & N. Anschuetz (Eds.), *The PDMA Handbook of New Product Development* (pp. 363-380). New York: Wiley.
35. Tian, X. (2013). Study on keyword extraction using word position weighted textrank. *Data Analysis and Knowledge Discovery*, 29(9), 30-34. Retrieved from [http://manu44.magtech.com.cn/Jwk\\_infotech\\_wk3/EN/10.11925/info-tech.1003-3513.2013.09.05](http://manu44.magtech.com.cn/Jwk_infotech_wk3/EN/10.11925/info-tech.1003-3513.2013.09.05)
36. Tsai, C.-Y., & Chiu, C.-C. (2004). A purchase-based market segmentation methodology. *Expert Systems with Applications*, 27(2), 265-276. <https://doi.org/10.1016/j.eswa.2004.02.005>
37. Wang, R., Zhang, W., Deng, H., Wang, N., Miao, Q., & Zhao, X. (2013). Discover community leader in social network with PageRank. In *Advances in Swarm Intelligence: 4th International Conference, ICSI 2013, Harbin, China*, 154-162. Berlin: Springer.
38. Wei, J. T., Lin, S. Y., Weng, C. C., & Wu, H. H. (2012). A case study of applying LRFM model in market segmentation of a children's dental clinic. *Expert Systems with Applications*, 39(5), 5529-5533. <https://doi.org/10.1016/j.eswa.2011.11.066>
39. Wei, J. T., Lin, S. -Y., Yang, Y. -Z., & Wu, H. -H. (2016). Applying data mining and RFM model to analyze customers' values of a veterinary hospital. *2016 International Symposium on Computer, Consumer and Control (IS3C)*, 481-484. <http://dx.doi.org/10.1109/IS3C.2016.126>
40. Wu, J., Shi, L., Lin, W. -P., Tsai, S. -B., Li, Y., Yang, L., & Xu, G. (2020). An empirical study on customer segmentation by purchase behaviors using a RFM model and K-means algorithm. *Mathematical Problems in Engineering*, 2020. <https://doi.org/10.1155/2020/8884227>
41. Yan, E., & Ding, Y. (2011). Discovering author impact: a PageRank perspective. *Information Processing & Management*, 47(1), 125-134. <https://doi.org/10.1016/j.ipm.2010.05.002>
42. Yoon, J.-L. (2007). Telco 2.0: A new role and business model. *IEEE Communications Magazine*, 45(1), 10-12. <http://dx.doi.org/10.1109/MCOM.2007.284530>
43. Zabkowski, T. S. (2016). RFM approach for telecom insolvency modeling. *Kybernetes*, 45(5), 815-827. Retrieved from <https://www.emerald.com/insight/content/doi/10.1108/K-04-2015-0113/full/html>
44. Zaheri, F., Farughi, H., Soltanpanah, H., Alaniazar, S., & Naseri, F. (2012). Using multiple criteria decision making models for ranking customers of bank network based on loyalty properties in weighted RFM model. *Management Science Letters*, 2(2), 697-704. <http://dx.doi.org/10.5267/j.msl.2012.01.018>